Patterns of molecular evolution and epistasis on a genomic and genic scale

Professor Daniel L. Hartl Pan-Pan Jiang

Patterns of molecular evolution and epistasis on a genomic and genic scale

Abstract

Epistasis describes non-additive interactions which affect gene expression and phenotype. It can happen on multiple levels, including on a genomic level with interactions between genes or even chromosomes affecting global patterns of gene expression. It can also happen within a gene itself, with epistatic interactions between amino acids affecting gene expression and resultant phenotypes. I present three studies in two organisms to study this phenomenon on a global-genomic scale, and also on a local-genic scale.

First, I present evidence that epistatic interactions between *Y*-linked regulatory polymorphisms and genetic background affect global gene expression in *Drosophila melanogaster*. The *Y* chromosome is a heterochromatic, degenerate chromosome and thought to have little evolutionary consequence. I studied *Y* chromosomes from two populations of *D. melanogaster* that are known to have major effects on the thermal tolerance of spermatogenesis. I show that these *Y* chromosomes differentially modify the expression of hundreds of autosomal and *X*-linked genes, but the effect depends on the genetic background the *Y* finds itself in. Second, I present novel evidence suggesting that the mechanism for Y-regulatory variation (YRV) is heterochromatin-based. Imprinting (due to parent-of-origin inheritance) in Drosophila has been documented mainly in heterochromatic regions, in particular the *Y* chromosome. I show that sex-specific transmission of the Y can lead to polymorphic imprinting and can change the magnitude

and scope of YRV, perhaps through differential titration of chromatin proteins. In particular, genes responding to this polymorphic imprint were more likely to be male-specific, testis-specific, and involved in rDNA transcript levels. This is particularly intriguing as rDNA processing is known to be affected by heterochromatin formation.

Finally, I study how mutational interactions within one gene can constrain evolutionary trajectories. The human malaria parasite, *Plasmodium vivax,* varies at several positions in the gene dihydrofolate reductase (DHFR), reflecting the mark of selection for drug resistance. Variation at four amino acid sites allows us to reconstruct a complete fitness landscape using all possible combinations of mutational variants within the gene. The results suggest that sign epistasis, where one mutation does well on some backgrounds but poorly on others, is common within *P. vivax* DHFR. In addition, drug concentration and effective population size can have a strong effect on whether the most resistant, quadruple-mutant, allele will fix in a population. I propose this may explain why the most resistant allele is missing from common polymorphic natural isolates.

Table of Contents

Preface

I would first like to thank Dan Hartl for the many ways in which he has positively shaped my life. First, his advice and insights into projects is unparalleled. Dan will always immediately understand what you're trying to explain, and even better, will quickly come up with innovative ways for solving particularly cumbersome experimental problems. Second, Dan is extremely supportive of diverse career goals, which is a rather rare trait in a PhD advisor. And lastly, Dan will do whatever is in his power to help you along your own life journey, whether that be writing thoughtful and persuasive recommendation letters, or offering to talk to you landlord when there is a tenancy dispute. He really goes above and beyond when it comes to his students, and for that, I am extremely thankful.

Just as it takes a village to raise a child, it takes a team to produce one PhD. Elena Lozovsky has been extremely gracious with her time, by helping me with experimental design and molecular techniques for my malaria project. Her expertise in microbiology is expansive, and I always know Elena will be able to answer any questions I have about cloning or primer design or yeast culture. Bernardo Lemos was instrumental in helping me design and implement my Y chromosome projects. I am particularly thankful for the countless hours he spent teaching me microarray techniques and best practices for data analysis. In addition, Bernardo is a great writer, and I have benefitted greatly from trying to emulate his scientific eloquence.

Next, I would like to thank my committee members, Hopi Hoekstra, Cassandra Extavour, and Anne Pringle, for their helpful comments and support along this process. Committee meetings, and even qualifying exams, are enjoyable with these committee members, who do their best to calm my nerves and always ask insightful questions.

In addition, I would like to thank the Hartl Lab, and in particular the following people, for their contribution to this thesis. Elaine Tran, a bright undergraduate, for her many hours of work culturing yeast and sending samples in for sequencing. I am sure Elaine will have a successful career in whatever field she chooses to go into. Hsiao-Han Chang for being my office-mate for almost 5 years, and for being there to share the ups and downs of graduate life in tandem. Russ Corbett-Detig for implementing the simulations I use for my malaria project, and for being patient and good teacher.

None of this would have been possible without my parents. My dad, for teaching me math and science since I was old enough to hold a pencil, which allowed me to impress people with my knowledge of the multiplication table at a very young age. My mom, for being supportive of all my career decisions and being generally super enthused about my research, no matter how esoteric it seemed. I would like to thank both my parents for cooking me delicious meals every time they visited Boston, which is even more impressive if you've seen the size of my kitchen in the first few years of graduate school.

Last but certainly not least, I would like to thank my partner in life, Mauricio, for his unconditional support, love, and computer-smarts. He has helped me solve seemingly

intractable computational problems. He has helped me see the bright side of less than stellar situations. And he has helped me define a career path I am excited about.

Chapter 1

Y not a dead end: Epistatic interactions between *Y*-linked regulatory polymorphisms and genetic background affect global gene expression in *Drosophila melanogaster*

Pan-Pan Jiang, Daniel L. Hartl, and Bernardo Lemos

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138

ABSTRACT

The *Y* chromosome, inherited without meiotic recombination from father to son, carries relatively few genes in most species. This is consistent with predictions from evolutionary theory that nonrecombining chromosomes lack variation and degenerate rapidly. However, recent work has suggested a dynamic role for the *Y* chromosome in gene regulation, a finding with important implications for spermatogenesis and male fitness. We studied *Y* chromosomes from two populations of *Drosophila melanogaster* that had previously been shown to have major effects on the thermal tolerance of spermatogenesis. We show that these *Y* chromosomes differentially modify the expression of hundreds of autosomal and *X*-linked genes. Genes showing *Y*-linked regulatory variation (YRV) also show an association with immune response and pheromone detection. Indeed, genes located proximal to the euchromatin-heterochromatin boundary of the *X* chromosome appear particularly responsive to *Y*-linked variation, including a substantial number of odorant-binding genes. Furthermore, the data show significant regulatory interactions between the *Y* chromosome and the genetic background of autosomes and *X* chromosome. Altogether, our findings support the view that inter-population, *Y*-linked regulatory polymorphisms can differentially modulate the expression of many genes important to male fitness, and they also point to complex interactions between the *Y* chromosome and genetic background affecting global gene expression.

INTRODUCTION

The *Y* chromosome is transmitted without sexual recombination from father to son. In the *Y* chromosome, as in other nonrecombining regions, complete linkage between genes results in the accumulation of deleterious alleles and the loss of genetic diversity due to the evolutionary processes of Muller's ratchet, background selection, and genetic hitchhiking (Bull, 1983; Rice, 1987, Charlesworth and Charlesworth, 2000; Bachtrog, et al. 2008). Consistent with theory, the *Y* chromosome of *Drosophila melanogaster* carries only thirteen known protein-coding genes (Carvalho et al 2001; Carvalho and Clark, 2005; Koerich et al 2008; Vibranovski et al. 2008; Krsticevic et al, 2010), whereas over 5,000 genes would be expected from typical gene densities in euchromatic regions.

Six of the thirteen genes discovered on the *Y* chromosome are male fertility factors that either encode structural components of spermatogenesis or regulate spermatogenesis-specific processes such as individualization (Carvalho 2000; Carvalho et al 2009). Spermatogenesis in Drosophila males is extremely sensitive to heat, with males becoming sterile anywhere from 23°C in heat-sensitive species to 31°C in heat-tolerant species (Chakir et al 2002; David et al. 2005). In *D. melanogaster*, Rohmer et al. (2004) found that differences between *Y*-chromosome lineages from tropical and temperate regions are responsible for much of the variation in thermal sensitivity of spermatogenesis. Since spermatogenesis is essential for male fitness, we expect a *Y*-chromosome effect on thermal sensitivity to translate into effects on male fitness. However, *Y*-chromosome effects on fitness can occur even at constant temperatures: Chippindale and Rice (2001) showed that polymorphisms in the *Y* chromosome have a large effect on male fitness, with total genetic variance in fitness comprising a limited contribution of additive genetic

variance but a substantial contribution of epistatic genetic variance. Accordingly, the contribution of a *Y* chromosome to fitness was found to be highly dependent on the genetic background of autosomes and the *X* chromosome.

While *Y*-linked protein coding genes show effectively no nucleotide diversity (average pair-wise differences, π) within *Drosophila melanogaster* (Zurovcova and Eanes, 1999) and very low levels of diversity in human populations (Shen et al., 1997; Rozen et al., 2009), *Y*-linked heterochromatic and rDNA repeats in both flies and humans can differ in repeat number or length (Karafet et al., 1998; Lyckegaard and Clark, 1989; Lyckegaard and Clark, 1991; Repping et al., 2003). Moreover, recent studies showed that the *Y* chromosome has undergone rapid evolution and turnover of protein-coding genes between humans and chimpanzee (Hughes et al., 2010) as well as among species of Drosophila (Koerich et al., 2008)

While no *Y*-linked transcription factors have been found in Drosophila, the *Y* chromosome is known to be a pervasive modulator of gene activity elsewhere in the genome. One phenomenon in which the *Y* chromosome affects expression of genes is position effect variegation (PEV) (Muller, 1930; Gatti and Pimpinelli, 1992; Talbert and Henikoff, 2006; Schulze and Wallrath, 2007). PEV occurs when genes are relocated next to a heterochromatin-euchromatin boundary. While these genes remain unchanged at the DNA level, they are transcriptionally repressed in some cells but not others. A classic example is the repositioning of the *w[m4]* allele from its normal location in euchromatin near the tip of the *X* chromosome to a new location close to an AT-rich microsatellite region in the pericentromeric heterochromatin near the base of the *X* (Muller, 1930). The variegated expression of *w[m4]* results in a mosaic, red-white eye-color phenotype. PEV-

associated repression of gene transcription is thought to result from the spread of pericentromeric heterochromatin into neighboring genes with consequent transcriptional silencing of those genes (Schulze and Wallrath, 2007). *Y*-chromosomal heterochromatin is known to suppress PEV in both *XY* males and *XXY* females (Dorer and Henikoff, 1994), with the level of suppression proportional to the amount of the *Y*-chromosome heterochromatin present (Dimitri and Pisano, 1989). One model for these effects is the competitive sequestration of chromatin-associated proteins by *Y*-linked microsatellite repeats (Lloyd et al., 1997; Wallrath, 1998).

Recent work by Lemos et al. (2008) has shown that the modulating effect of cryptic *Y*-chromosome polymorphisms on gene expression is pervasive throughout the *D. melanogaster* genome. Accordingly, males differing only in the origin of their *Y* chromosome showed differential expression at hundreds of non-*Y*-linked genes. Interestingly, many of these genes have male-biased expression and seem to be involved in species divergence and temperature adaptation. These results provided a molecular framework for how the *Y* chromosome affects adaptive phenotypic variation and fitness (Voelker and Kojima, 1971; Chippindale and Rice, 2001; Rohmer et al., 2004).

The role of genetic background on *Y*-linked regulatory variation (YRV) remains to be addressed. Previous experiments by Lemos et al. (2008) placed *Y* chromosomes in the genetic background of an inbred, homogeneous laboratory strain (B4361). This laboratory strain was chosen to ensure uniformity of genetic background. However, autosomal and *X*-chromosome polymorphisms occurring in natural populations may lend themselves to subtle modifications by cryptic *Y*-linked regulatory polymorphisms. In accordance with this possibility, Chippindale and Rice (2001) found significant

background-by-*Y* interaction effects on male fitness. In addition, no studies have yet investigated the physical clustering of the genes affected by YRV along chromosomes.

This study is aimed at addressing the following questions: (i) Which genes show regulatory modulation due to *Y*-by-background effects? (ii) Which functional categories do these genes fall into? (iii) Do the genes affected by YRV show distinctive physical clustering patterns along autosomes or *X* chromosome? (iv) Is PEV differentially modulated by specific Y-by-background combinations? The *Y* chromosomes chosen were sampled from a tropical (India) and temperate (France) population of *D. melanogaster*. Flies from these populations have previously been shown to have major differences in their ability to carry out spermatogenesis under heat stress, in large part due to polymorphic variation between their *Y* chromosomes (Rohmer et al. 2004). Here we test the effect of the *Y* chromosomes on gene expression not only in the genetic background of an inbred laboratory strain, but also in the genetic background of both the tropical and temperate populations from which the *Y* chromosomes were derived. This experimental design allowed us to address the extant to which the expression of polymorphic *Y*-linked variation depends on the subtleties of genomic background. A gene-density plotting algorithm was used to test for physical clustering of genes showing YRV. Finally, both naturally occurring *Y* chromosomes were assayed for polymorphisms capable of regulating PEV.

MATERIALS AND METHODS

Fly strains:

Wild flies were collected from Draveil, France (in 2001) and Delhi, India (in 1997) by members of Jean R. David's research group (Rohmer et al, 2004), and generously provided to us for analysis. For each population, wild-collected females were isolated in culture vials, and isofemale lines established. These lines (a minimum of 10) were eventually pooled to make a mass culture. Approximately 100 males per population were drawn from mass culture for crosses (Figure S1.1) which allowed *Y* chromosomes from the French population to be introgressed into an otherwise Indian autosomal and *X*-chromosomal background, and vice versa. *Y* chromosomes from both populations were also introgressed into the same laboratory genetic background (B4361) used by Lemos et al. (2008). Hence, six strains were used for analysis. These consisted of the original French and Indian lines with their native *Y* chromosomes, namely: French background with French *Y* chromosome (French : Y_F), and Indian background with Indian *Y* chromosome (Indian : Y_I). In addition we studied four *Y*-substituted strains, namely, Indian genetic background with French *Y* chromosome (Indian : Y_F), laboratory background with French *Y* chromosome (B4361 : Y_F), France background with Indian *Y* chromosome (French : Y_I), and laboratory background with Indian *Y* chromosome (B4361 : Y_I). Males from these strains were collected for use in microarray dye-swap experiments. Newly emerged males were collected on the tenth day after egg laying and allowed to age for three days at 25°C, after which they were flash frozen in liquid nitrogen and stored at -80°C.

All crosses for each *Y*-substitution line were carried out with 10–15 vials with multiple mating pairs per vial. Gene expression differences between males were assayed in flies grown under carefully controlled environments - 24h light, 25°C, and constant

humidity - and harboring naturally occurring Y-chromosome variants.

Microarray hybridizations and analysis:

Our experimental design consisted of 16 cDNA microarrays, 4–6 for each of the three backgrounds (French, Indian, and B4361), involving 32 separate labeling reactions. We contrasted two *Y* chromosomes (Y_I and Y_F) on each microarray. Microarrays were ~18,000-feature cDNA arrays spotted with *D. melanogaster* cDNA PCR products as described (Lemos et al., 2008). RNA extraction, cDNA synthesis, microarray hybridization, and microarray slide scanning protocols closely followed that of Lemos et al (2008). Foreground fluorescence of dye intensities was normalized by the Loess method implemented in the library Limma of the statistical software R. Microarray gene expression data herein reported can be obtained at the GEO database (GSE21587).

Significance of variation in gene expression in each background due only to the *Y* chromosome was assessed using the Bayesian Analysis of Gene Expression Levels (BAGEL) algorithm (Townsend and Hartl, 2002). False discovery rates (FDR) were estimated empirically based on the variation observed when randomized versions of the original dataset were analyzed. Density along chromosomes of genes showing *Y*-linked regulatory variation as assessed by BAGEL was plotted using a sliding window algorithm with window size of 2 Mb, sliding in 1 Mb increments. Confidence intervals were estimated empirically by running the density-plotting algorithm on 1000 sets of randomly sampled genes taken from the genome as a whole as represented by features on the microarray, with gene number equal to the number of differentially expressed genes.

Cutoff 95% densities were plotted, and any clusters with observed densities beyond these values were regarded as significant.

To test for the effect of *Y*-by-background interaction on gene expression, a linear model was fitted to normalized data: $\gamma_{ij} = \mu + B_i + Y_j + \text{I(Background} \times Y)_{ij} + e_{ijk}$, where γ_{ij} is the normalized-transformed gene expression, μ is the population mean, B_i is the effect of the i^{th} genetic background, Y_j is the effect of the j^{th} *Y* chromosome, $\text{I(Background} \times Y)_{ij}$ is the effect of background-by-*Y* interaction, and e_{ijk} is the residual effect. To test for agreement with the BAGEL results, a second model was implemented to test for Y-only effects: $\gamma_{ij} = \mu + Y_j + e_{ijk}$, where Y_j refers to Y_I or Y_F. The significance of effects from *Y*-only and background-by-*Y* interactions were tested using the *Fs*-test, a modified *F*-statistic incorporating shrinkage variance components (Cui et al., 2005). *P* values were calculated by performing 1000 permutations of samples, and corrected for multiple hypotheses testing by the *q*-value false discovery rate method (Storey and Tibshirani, 2003). Significant changes were determined at the FDR threshold of 0.01. A *k*-means analysis was used to identify groups of genes with similar expression patterns across *Y*-by-background groups. In the bootstrapped *k*-means algorithm, a gene was assigned to a group if it was identified in 80% of 1,000 iterations. This was repeated for different values of *k* to find the *k* needed to minimize the number of genes not identified in any group. All analyses described in this paragraph were computed with the R/Maanova package (Wu et al., 2003)

Enrichment in gene ontology categories was assessed with GeneMerge (Castillo-Davis and Hartl, 2003), which uses a hypergeometric distribution to assess significance. Because GeneMerge tests for all categories, a modified Bonferonni correction was used

to account for multiple testing.

Position effect variegation (PEV):

Males from all four populations were crossed to females from a stock carrying *w[m4h]* (Figure S2). These females possess an inversion on the *X* chromosome that repositions the *w[m4h]* gene proximal to the *X*-centromere. All flies were maintained at either 25°C or 18°C. Males from these crosses were collected, flash frozen in liquid nitrogen, aged for 3 days at either 25°C or 18°C, and stored at -80°C. Heads of males were removed with a blade, and homogenized 5 to a tube with 10uL of acidified ethanol (30% ethanol acidified to pH 2 with HCl). Eye pigment expression was assessed with spectrophotometric analysis at an optical density of 480 nm. 4–6 biological replicates were used per treatment, with two measurements taken per replicate. The correlation between repeat measures was high (Pearson's $r = 0.90$), thus their means were used in subsequent analyses. Males displaying typical eye-pigmentation phenotypes were imaged using an auto-montage system (Snycroscopy, Frederick, MD). A 3-way ANOVA analysis, using statistical software JMP, was performed using male background (Indian, French, or B4361), *Y* chromosome (Y_I or Y_F), and temperature (25°C or 18°C) as factors.

RESULTS

Global gene expression variation:

Males differing only in their *Y* chromosomes (either Y_I or Y_F) showed differential expression of a substantial number of genes, with the exact number depending on the genetic background interrogated and the cutoff for significance used (12 to 1178 genes

when the Bayesian posterior probability is > 0.999 or 0.90, corresponding to false discovery rates [FDR] of < 1% or 35%, respectively; Figure 1.1]. For every genetic background, and at every significance cutoff value, the observed number of genes differentially expressed among Y_I and Y_F males exceeded the number expected by chance. Overlap of differentially expressed genes sets between the three different backgrounds is shown in Figure 1.2. The results suggest that although *Y*-linked regulatory polymorphisms have different modulating effects on genes depending on the genetic background of the male, there is some agreement on the genes showing *Y*-linked regulatory variation (YRV). All the following analyses are based on genes affected by YRV identified with a criterion of a Bayesian posterior probability > 0.95.

Clustering of genes showing YRV into significant functional categories is listed in Table 1.1. Of note, pheromone binding and immune response genes are heavily represented, as well as genes localized to extracellular regions across all three backgrounds. This again suggests that genes showing YRV show a consistent pattern with respect to function.

Physical clustering of genes along chromosomes was also examined. While no physical clustering is apparent in the autosomes, males possessing Y_F showed overexpression of genes near the euchromatin-heterochromatin boundary of the *X* chromosome (at chromosome position 22 Mb) as compared to males possessing Y_I (Figure 1.3). This pattern holds true in both the Indian and French genetic backgrounds, but not in the B4361 genetic background. In the Indian background, 5 genes near the *X*-chromosome euchromatin-heterochromatin boundary showed this pattern of overexpression in Y_F males compared to Y_I. Of these, one encodes a protein categorized

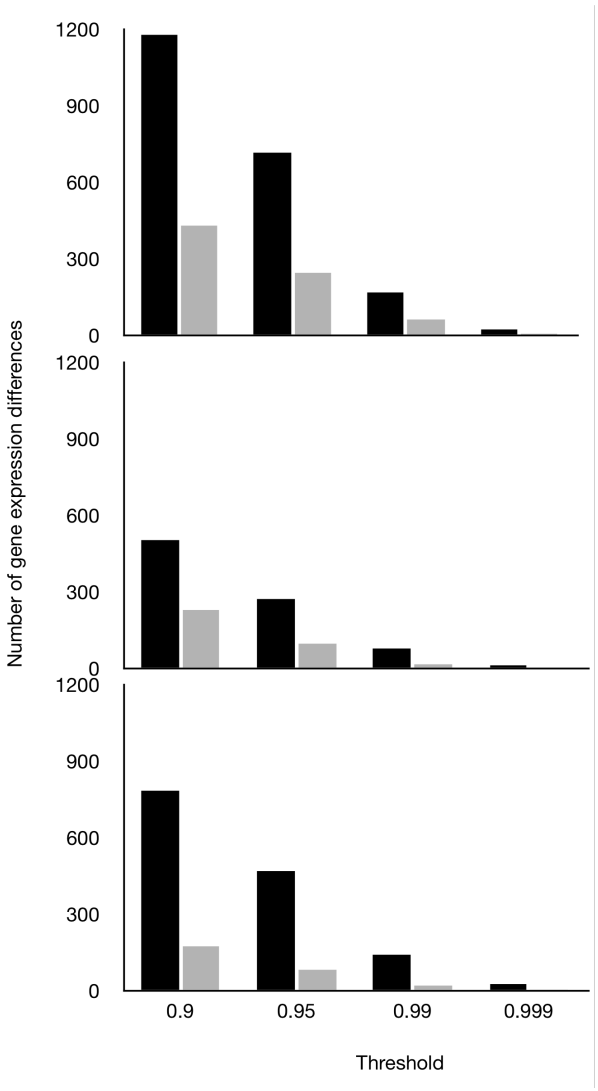


Figure 1.1 – Black bars represent the number of genes differentially expressed by males possessing Y_I or Y_F in an Indian genetic background (top panel), French genetic background (middle panel) or B4361 genetic background (lower panel), as a function of the Bayesian posterior probability of differential expression. Gray bars indicate the estimated number of genes expected by chance.

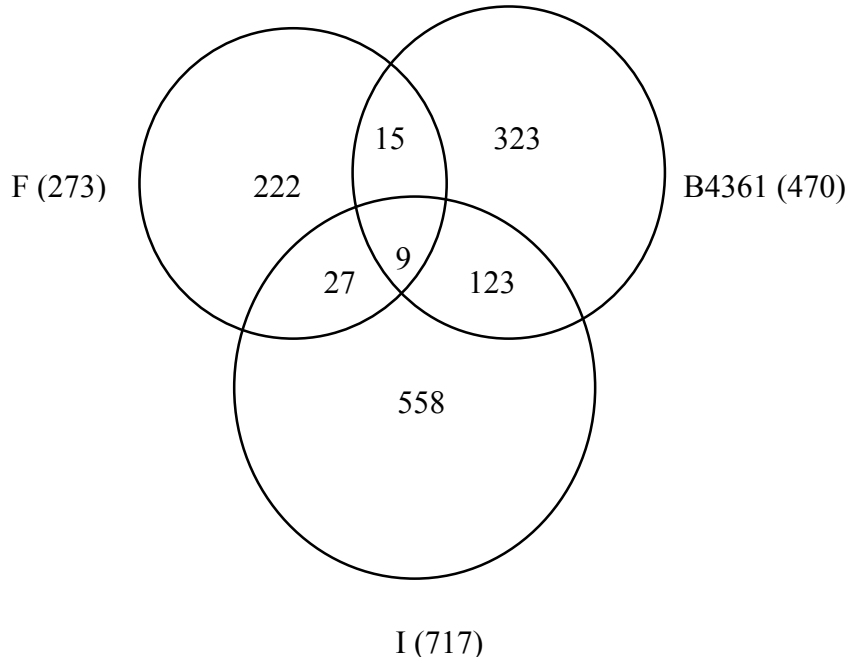


Figure 1.2 – Venn diagrams showing number of unique *Y*-regulated genes (Bayesian posterior probability > 0.95) in Indian (I), French (F), and B4361 genetic backgrounds, and overlap between affected genes in different backgrounds. Number of genes found to be significant in each individual background is given in parentheses.

Table 1.1 – Significantly over-represented Gene Ontology categories for genes identified to be overexpressed in BAGEL analysis. *P*-values are adjusted for multiple hypothesis testing.

		Background		
		I	F	B4361
GO category	Description	Number of Genes		
Molecular Function				
GO:0004252	serine-type endopeptidase activity	14*	4	3
GO:0005550	pheromone binding	5	6***	1
GO:0008145	phenylalkylamine binding	2	3*	1
GO:0004364	glutathione transferase activity	10	7*	1
Biological Processes				
GO:0006952	defense response	47***	21*	17
GO:0009636	response to toxin	19**	11**	4
GO:0006508	proteolysis	62***	18	24
GO:0006961	antibacterial humoral response	5	6*	7
GO:0019236	response to pheromone	2	4*	2
Cellular Component				
GO:0005576	extracellular region	34***	19***	25*

* $P < 0.05$

** $P < 0.01$

*** $P < 0.001$

Figure 1.3 – Clustering of genes showing YRV along the euchromatic portion of the *X* chromosome in an Indian genetic background (top panel), French genetic background (middle panel) or B4361 genetic background (lower panel). Black lines indicate observed density of genes around a 2 Mb sliding window (step size 1 Mb). Grey lines indicate 95% confidence intervals. A schematic *X* chromosome is drawn in the top left panel with light areas representing euchromatin and dark areas representing heterochromatin. The dark knob at the right represents the X chromosome centromere. Columns represent genes for which males possessing Y_F showed enhanced expression over males possessing Y_I ($Y_F > Y_I$) or vice versa ($Y_I > Y_F$). An asterisk denotes a chromosomal segment containing significantly more genes showing *Y*-linked regulation than expected by chance.

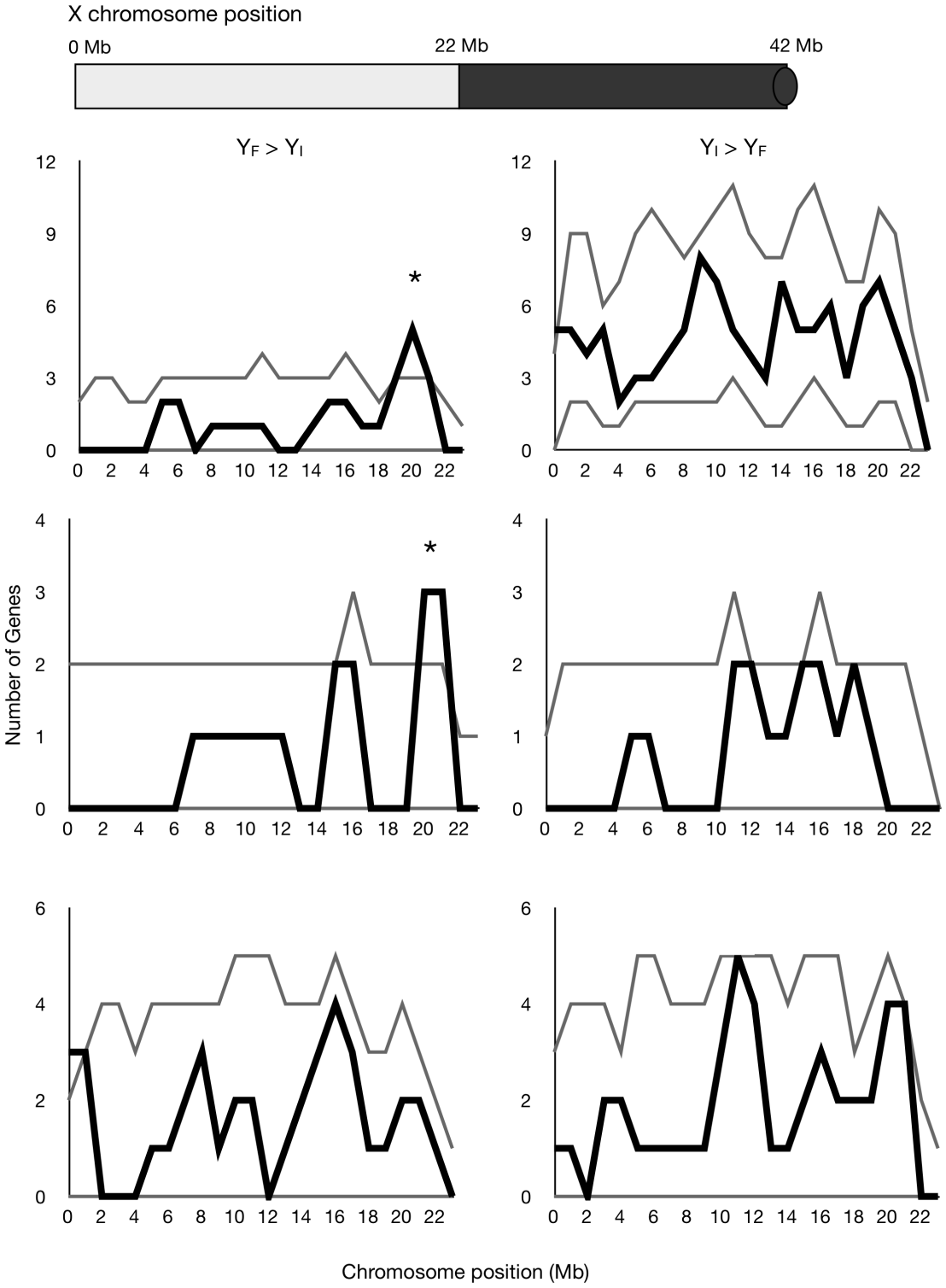


Figure 1.3 (Continued)

as sensory perception of chemical stimulus, namely gene *Obp19a* encoding odorant-binding protein 19a. In the French genetic background, 3 genes near the *X*-chromosome euchromatin-heterochromatin boundary showed this pattern. All three encode proteins involved in the sensory perception of chemical stimulus (*Obp19a* and *Obp19b* encoding odorant-binding proteins 19a and 19b, and *Pbprp3* encoding pheromone-binding protein-related protein 3). The observed effect of the *Y* chromosome on modulating genes associated with pheromone binding and sensory perception has important implications for male fitness and sexual selection.

The relative expression levels of three representative genes showing YRV are plotted in Figure 1.4. One gene, *Obp19b*, was chosen because it is located near the *X*-chromosome euchromatin-heterochromatin boundary and was also identified in both the French and Indian backgrounds as being significantly overexpressed in Y_F compared to Y_I males. Two other genes, *Dro* (Drosocin) and *Pbprp3* (pheromone-binding protein-related protein 3) were chosen because they belonged to biological clusters identified to be over-represented in the YRV-regulated gene sets (immune response and pheromone binding, respectively). Ratios represent relative expression levels of genes in Y_F males over Y_I males in the three genetic backgrounds. The results show, again, how epistatic interactions between *Y*-linked polymorphisms and background can modulate the expression of non-*Y*-linked genes. Particularly striking is the differential relative expression of *Dro,* an immune response gene, in the three backgrounds: when comparing expression of *Dro* in Y_F males to Y_I males, it is underexpressed in the Indian background,

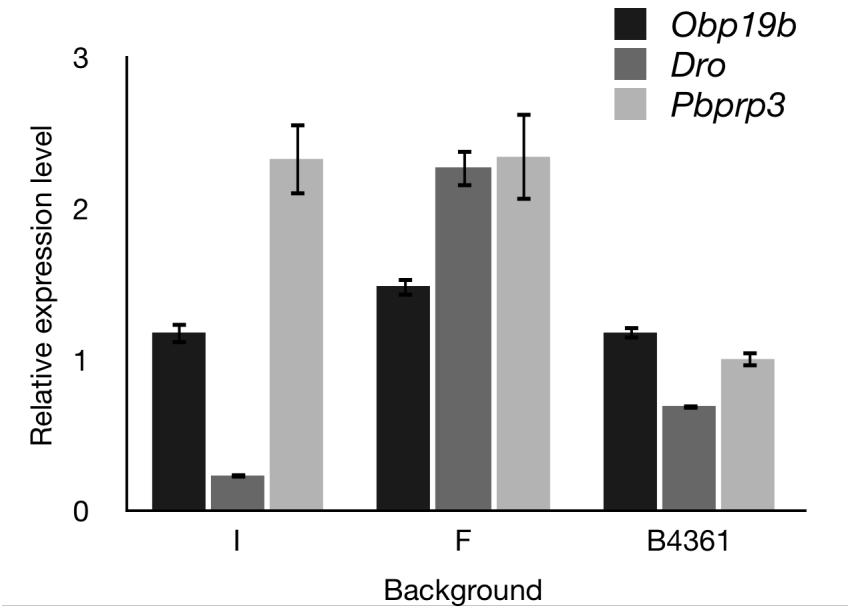


Figure 1.4 – Relative expression levels of three genes showing YRV in Y_F males versus Y_I males in three genetic backgrounds (I, Indian; F, French; B4361). Expression levels are shown as the ratio of Y_F over Y_I expression (± SE). *Obp19a* is an odorant binding-protein gene near the *X* euchromatin-heterochromatin boundary. Drosocin (*Dro*) is an immune-response gene. *Pbprp3* encodes pheromone-binding protein-related protein 3, a pheromone-binding protein.

overexpressed in the French background, and slightly underexpressed in the B4361 background.

Extensive Y-by-background interaction:

Y–by-background effects, as assessed by the Maanova linear model, influenced the expression of 346 genes (FDR < 0.01). Agreement between these results and the BAGEL results is strong: 252 (74%) of the 346 genes were also identified by BAGEL as differentially regulated by *Y*-linked polymorphisms in at least one of the genetic backgrounds (4361, Indian, or French) (Table S1.1). 200 (57.8%) of the 346 genes can be grouped most parsimoniously into three clusters of gene expression (Figure 1.5). In each cluster, genes show similar patterns of high expression in some *Y*-background combinations and low expression in others. Significant Gene Ontology categories within each cluster are listed in Table 1.2. In both methods of analysis (BAGEL and Maanova), immune response genes are heavily represented within significantly differentially expressed genes.

Y-chromosome main effects, as identified by Maanova, regulate the differential expression of 192 genes. A proper comparison with BAGEL results called for re-analysis of global gene expression patterns across all arrays (regardless of genetic background) using BAGEL. When this was done, 484 genes showed differential gene expression (Bayesian Posterior Probability > 0.95). Of the 192 genes identified by Maanova, 106 (55%) of them were also identified in the BAGEL gene set, while 86 (45%) were not. The 86 novel targets of YRV are not surprising in view of our results indicating strong

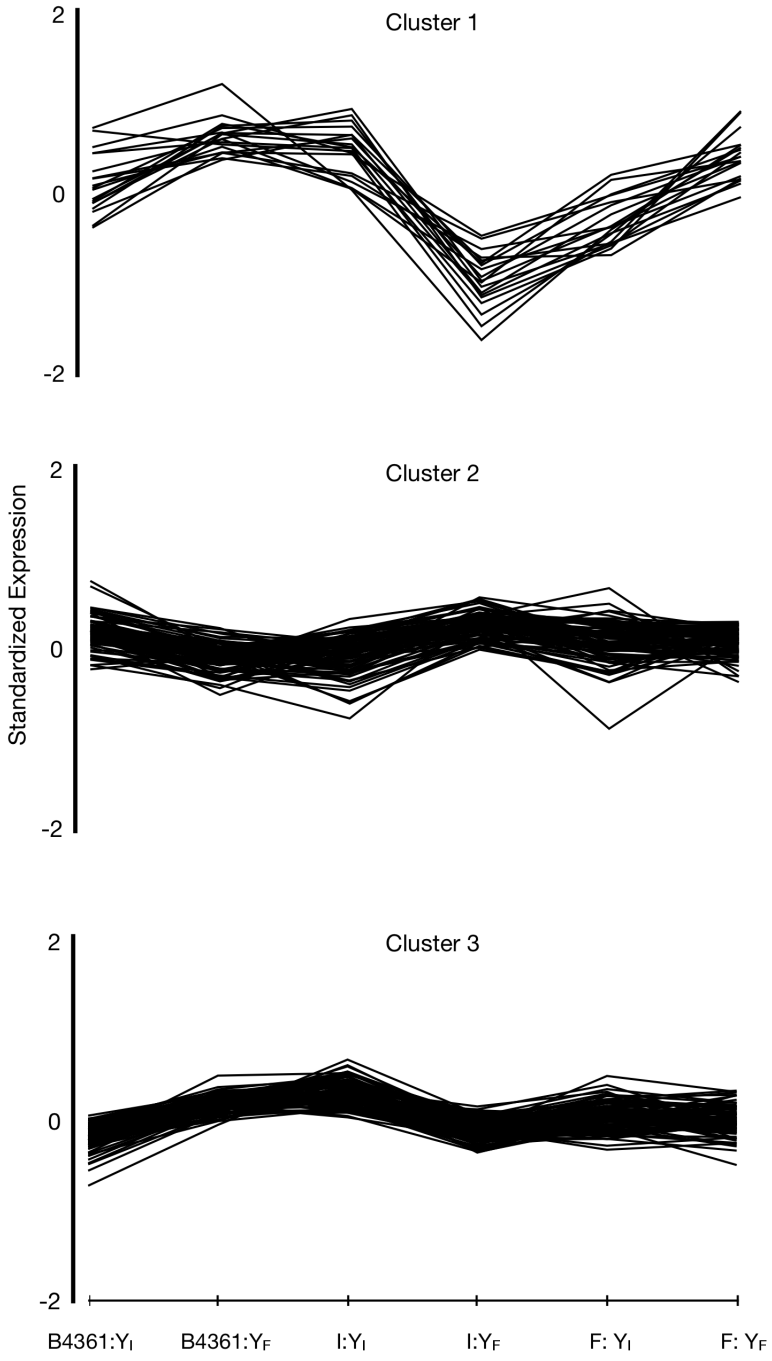


Figure 1.5 – Gene expression profiles generated by *k*-means clustering. Each line represents the expression of one gene across each background-by-*Y* group. Expression measures were standardized across groups, and analyzed using *k*-means cluster analysis. There are 19, 88, and 93 genes in clusters 1, 2, and 3, respectively.

Table 1.2 – Significantly over-represented functional Gene Ontology categories for gene expression clusters as identified by Maanova for *Y*-by-background effects. *P*-values are adjusted for multiple hypothesis testing.

Cluster	GO category	Description	Number of genes
1	GO:0006961	antibacterial humoral response (sensu Protostomia)	6***
1	GO:0042742	defense response to bacterium	4***
1	GO:0050829	defense response to Gram-negative bacterium	3***
1	GO:0050830	defense response to Gram-positive bacterium	2**
1	GO:0001501	skeletal development	2**
1	GO:0005975	carbohydrate metabolic process	3*
2	GO:0019236	response to pheromone	3**
2	GO:0045861	negative regulation of proteolysis	2*
3	GO:0009405	pathogenesis	2*
3	GO:0006631	fatty acid metabolic process	4*

* $P < 0.05$

** $P < 0.01$

*** $P < 0.001$

epistatic *Y*-by-background effects on gene expression. Such genes can fail to be identified by BAGEL because genes that are highly modulated by the *Y* chromosome in one genetic background but less so in others might show no consistent difference in expression between the temperate and tropical *Y* chromosomes when averaged across all genetic

backgrounds.

Polymorphic Y chromosome effects on position effect variegation:

The above results from microarray data suggesting large *Y*-by-background interaction effects on gene expression, as well as clustering of these effects near the *X*-chromosome euchromatin-heterochromatin boundary, were confirmed with a PEV assay. Males in the assay possessed either a Y_I or Y_F in a hybrid genetic background consisting of white-eye mutation *w[m4h]* positioned near the X chromosome euchromatin-heterochromatin boundary, a haploid autosomal genome sampled from the original stock containing the PEV marker, and a haploid autosomal genome sampled from the Indian, French, or B4361 laboratory populations. The results suggest that temperature does not affect suppression of PEV ($P = 0.26$). This result is surprising, as previous studies have shown that high temperatures during development suppress PEV, while low temperatures enhance PEV (Spofford, 1976; Zhang and Stankiewicz, 1998). There were also no significant temperature-interaction factors (temperature × *Y* chromosome, $P = 0.92$; temperature × genetic background, $P = 0.41$; temperature × background × *Y* chromosome, $P = 0.73$). On the other hand, Y_I and Y_F differed dramatically in their effects on position effect variegation (Figure 1.6, $P < 0.0001$), with Y_I males showing broader expression of *w[m4h]* than Y_F in all genetic backgrounds; however the effect is least pronounced in the

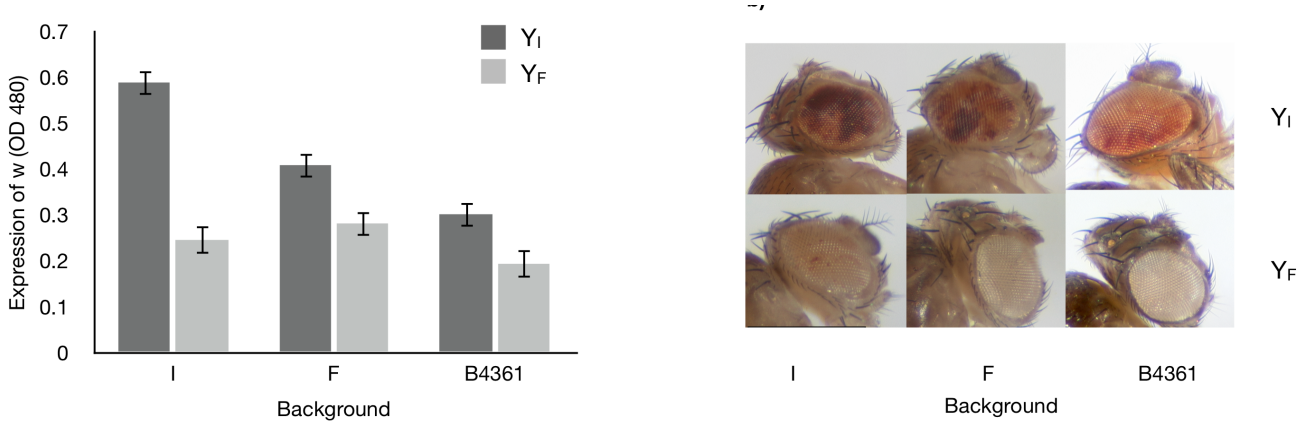


FIGURE 1.6. – *Y*-chromosome effects on position effect variegation (PEV). Y_I suppresses PEV more, and thus allows more expression of *w[m4h]*, than Y_F in all three genetic backgrounds (I, Indian; F, French; B4361). Eye pigmentation was measured as absorption of light at 480 nm. Pictures of heads of representative male flies are shown to the right.

B4361 background. Also importantly, genetic background showed a significant effect on PEV ($P < 0.0001$), with a similarly significant effect for *Y*-by-background interaction ($P < 0.001$). These results suggest that the modulation of PEV, which is caused by the mosaic expression of *w[m4h]*, a gene positioned near the X chromosome euchromatin-heterochromatin boundary, is sensitive to epistatic interactions between the *Y* chromosome and genetic background. These results are therefore in agreement with the findings from our genome-wide gene expression assay.

DISCUSSION

The data presented here suggest that polymorphic variation in *Y* chromosomes from two geographically diverse *D. melanogaster* populations differentially regulate the

expression of hundreds of autosomal and *X*-linked genes. However, the contribution of *Y* chromosomes to global expression profiles depends on the genetic background of the bearer. Accordingly, we observed that the contribution of a temperate or a tropical *Y* chromosome to global gene expression is most pronounced when males differing only in the origin of their Y-chromosomes are assayed in their wild-type naturally occurring genetic backgrounds: Y-chromosome substitution lines in the Indian background showed more than twice the number of differentially expressed genes than Y-chromosome substitution lines in the French background.. This study also presents new data suggesting the physical clustering of genes exhibiting *Y*-linked regulatory variation (YRV). We found significant physical and functional clustering around the euchromatin-heterochromatin boundary of the *X* chromosome, with *X*-linked olfaction-related genes showing higher transcription levels in males with Y_F than in males with Y_I. Finally, the *Y*-by-background interaction effects on autosomal and *X*-linked gene expression, as well as the existence of polymorphic variation between the two *Y* chromosomes in their effects on modulating genes proximal to the euchromatin-heterochromatin boundary in the *X*-chromosome, were confirmed with a position-effect variegation assay.

Y-linked genetic variation has been previously documented for sex ratio (Carvalho et al., 1997; Montchamp-Moreau et al., 2001), male courtship (Huttunen and Aspi, 2003); geotaxis (Stoltenberg and Hirsch, 1997), thermal sensitivity of spermatogenesis (Rohmer et al., 2004), and fitness (Chippindale and Rice, 2001). For many of these traits (male courtship, geotaxis, spermatogenesis, and fitness), significant *Y*-by-background interaction effects have also been detected. Thus, our observations regarding substantial

Y-by-background interaction for gene expression traits are in good agreement with these previous findings regarding higher-level phenotypes.

These findings of *Y*-chromosome effects on male phenotypes contrast with molecular analyses showing no polymorphism among 11 alleles of a 1738-bp region of a *Y*-linked gene in *D. melanogaster* (Zurovcova and Eanes, 1999). In humans, *Y*-linked genes also show decreased levels of molecular variation, with a large-scale analysis of four *Y*-linked genes finding that coding regions show between 0% and 20% of the polymorphism of a sample of autosomal genes (Shen et al., 1997; Rozen et al., 2009). Despite the lack of nucleotide diversity in coding sequences of *Y*-linked genes, considerable structural polymorphism has been detected in the copy numbers of *Y*-linked heterochromatin repeats in humans and flies (Karafet et al., 1998; Lyckegaard and Clark, 1989; Lyckegaard and Clark, 1991; Repping et al., 2003). Repeat sites have been shown to act as nucleation sites for heterochromatin formation via the RNAi pathway (Dorer and Henikoff, 1994; Elgin and Grewal, 2003; Volpe et al., 2002; Pal-Bhadra et al., 2004).

Heterochromatin can influence transcription epigenetically, with the effect most easily observed in the modification of PEV by the *Y* chromosome (Dimitri and Pisano, 1989; Dorer and Henikoff 1994). Large heterochromatic blocks, such as the *Y* chromosome, are thought to sequester limiting heterochromatin factors from other regions, thus impeding the spread of heterochromatin to nearby loci (Lloyd et al., 1996; Schulze and Wallrath, 2007). In this way, silencing of genes located near heterochromatin-euchromatin boundaries is suppressed. Balanced polymorphisms in *Y*-chromosome heterochromatin repeats may provide the necessary molecular variation for differential competitive binding ability of chromatin proteins. The concomitant

redistribution of chromatin proteins will most strongly influence the expression of genes located next to other heterochromatic blocks, such as at euchromatin-heterochromatin boundaries. An alternate explanation for the influence of the *Y* chromosome on PEV is via the *Y* chromosome's effect on RNA-interference pathways. The spread of heterochromatin is initiated through the transcription of repeat-DNA, and then propagated via the RNA-interference pathway. Lemos et al. (2008) found *Y*-linked polymorphisms responsible for the differential expression of transposable elements, which are known to undergo RNAi-mediated silencing. Therefore, mechanistic similarities underlying *Y*-linked effects on gene expression and PEV may exist.

The acrocentric X chromosome of *D. melanogaster* is partitioned into 22 Mb of distal euchromatin, and 20 Mb of proximal heterochromatin (Adams et al 2000). As our results suggest, many of the genes showing *Y*-linked regulatory variation near the euchromatin-heterochromatin boundary in the *X*-chromosome are odorant-binding proteins, which are components of the insect olfactory system (Wang et al., 2009). Odorant receptors are rapidly evolving molecules in the Drosophila proteome (Robertson et al. 2003) and show altered expression following mating (McGraw et al. 2004). Interestingly, the genes affected by *Y*-linked regulatory elements exhibit significant functional coherence in showing association with pheromone detection. The influence of the *Y* chromosome on pheromone detection as well as odorant-binding proteins suggests a role for the *Y* chromosome in mating behavior and may help to explain the cessation of rigorous courtship and the reduced mating success of *X*O *Drosophila* males (Cordts and Partridge, 1996; Kuijper et al., 2006). In Anopheles mosquitoes, the *Y* chromosome has also been implicated in influencing mating behavior (Fraccaro et al., 1977). Because

many mating-behavior related proteins are selected for in different ways in males and females, there may be selection for sex-limitation of modifiers of their expression. Since the *Y* chromosome is male-limited, it serves as the perfect platform for these modifiers. Lemos et al. (2008) showed that genes showing *Y*-linked regulatory variation are more highly expressed in males than females, suggesting that the recruitment of modifiers of male-biased genes may have shaped the evolution of the *Y* chromosome.

In addition to pheromone binding proteins, genes showing *Y*-linked regulatory variation are also associated with immune response and are more likely to be localized to the extracellular matrix than expected by chance. Although there are no previous studies of *Y* chromosome effects on immune response genes in *Drosophila*, studies in mice have found that *Y*-linked polymorphisms are capable of modifying autoimmune disease susceptibility (Teuscher et al., 2006; Spach et al., 2009). However, in mice, several genes of immunologic significance are located on the *Y* and may serve as candidates for explaining the effect. In *Drosophila*, no *Y*-linked immune-related genes are known. Therefore we suggest that our findings of *Drosophila* immune response genes being responsive to YRV are most likely explained by variation in non-coding components of the *Y* chromosome, such as repeat copy number. *D. melanogaster* populations from France and India are known to differ in the thermal sensitivity of spermatogenesis (Rohmer et al. 2004, David et al. 2005), with temperate and tropical *Y*-chromosomes contributing substantially to this difference. Among genes that show YRV in at least two of the three backgrounds, we find candidates known to be structural constituents of cytoskeleton (*nod*, *CG9279*, and *tm2*) and lipid metabolism (*CG9914*, *CG17292, CG9458, CG11426, CG6295, CG6277, CG18815, CG31872*). In addition, fatty acid

metabolism genes are overrepresented in one of the clusters detected by *k*-means analysis using Maanova. This suggests that, while the heat sensitivity of spermatogenesis express itself sharply at higher temperatures, the modulating effects of the *Y* chromosome on sperm-related traits may be subtle at permissive or less stressful temperatures. Lastly, localization of genes showing YRV to extracellular regions is expected, as many pheromone-binding proteins and proteins involved in immune response are receptor proteins with large extracellular domains.

In summary, our finding of cryptic *Y*-linked regulatory control of hundreds of genes across various genetic backgrounds suggests standing *Y*-linked balanced polymorphisms in natural populations. At a cursory glance, this result seems incongruent with previous theoretical and empirical work suggesting little *Y*-linked polymorphism can be supported in a nonrecombining chromosome. However, our findings, together with other studies of the nontransitivity of sperm competition and *Y*-by-background interactions for male fitness (Clark et al., 2000; Chippindale and Rice, 2001), bring to light some complex and previously underappreciated dynamics for maintaining *Y*-linked polymorphisms.

LITERATURE CITED

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne *et al*, 2000 The genome sequence of drosophila melanogaster. Science **287:** 2185-2195.

Bachtrog, D., E. Hom, K. M. Wong, X. Maside, and P. de Jong. 2008 Genomic degradation of a young *Y* chromosome in *Drosophila miranda*. Genome Biology **9**: R30.

Bull, J.J. 1983 *Evolution of Sex Determination Mechanisms*. Benjamin Cummings, Menlo Park, CA.

Carvalho, A. B., S. C. Vaz, and L. B. Klaczko. 1997 Polymorphism for *Y*-linked suppressors of sex-ratio in two natural populations of *Drosophila mediopunctata*. Genetics **146**: 891-902.

Carvalho, A. B., B. P. Lazzaro, and A. G. Clark. 2000 Y chromosomal fertility factors kl-2 and kl-3 of *Drosophila melanogaster* encode dynein heavy chain polypeptides. Proc. Natl. Acad. Sci. USA **97:** 13239-13244.

Carvalho, A. B., B. A. Dobo, M. D. Vibranovski, and A. G. Clark. 2001 Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **98**: 13225-30.

Carvalho, A. B. and A. G. Clark. 2005 Y chromosome of *D. pseudoobscura* is not homologous to the ancestral Drosophila Y. Science **307**:108-10.

Carvalho, A. B., L. B. Koerich, and A. G. Clark. 2009 Origin and evolution of Y chromosomes: Drosophila tales. Trends Genet. **25**:270-7.

Castillo-Davis, C. I. , and D. L. Hartl. 2003 GeneMerge--post-genomic analysis, data mining, and hypothesis testing. Bioinformatics **19:** 891-892.

Chakir M., A. Chafik, B. Moreteau, P. Gibert, and J. R. David. 2002 (Weinreich et al. 2006; Lozovsky et al. 2009; Brown et al. 2010; Chou et al. 2011) Genetica **114**:195-205.

Charlesworth, B., and D. Charlesworth. 2000 The degeneration of *Y* chromosomes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **355:** 1563-1572.

Chippindale, A. K., and W. R. Rice. 2001 Y chromosome polymorphism is a strong determinant of male fitness in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **98:** 5677-5682.

Clark, A. G., E. T. Dermitzakis, and A. Civetta. 2000 Nontransitivity of sperm precedence in Drosophila. Evolution **54:** 1030-1035.

Cordts, R., and L. Partridge. 1996 Courtship reduces longevity of male *Drosophila melanogaster*. Animal Behaviour **52:** 269-278.

Cui, X., J. T. Hwang, J. Qiu, N. J. Blades, and G. A. Churchill. 2005 Improved statistical tests for differential gene expression by shrinking variance components estimates. Biostatistics **6:** 59-75.

David, J. R., L.O. Araripe, M. Chakir, H. Legout, B. Lemos, *et al*. 2005 Male sterility at extreme temperatures: a significant but neglected phenomenon for understanding Drosophila climatic adaptations. Journal of Evolutionary Biology **18:** 838-846.

Dimitri, P., and C. Pisano. 1989 Position effect variegation in *Drosophila melanogaster*: relationship between suppression effect and the amount of *Y* chromosome. Genetics **122:** 793-800.

Dorer, D. R., and S. Henikoff. 1994 Expansions of transgene repeats cause heterochromatin formation and gene silencing in drosophila. Cell **77:** 993-1002.

Elgin, S. C. R., and S. I. S. Grewal. 2003 Heterochromatin: silence is golden. Current Biology **13:** R895-R898.

Fraccaro, M., L. Tiepolo, U. Laudani, A. Marchi, and S. D. Jayakar, S.D. 1977 *Y* chromosome controls mating behaviour in Anopheles mosquitos. Nature **265:** 326-328.

Gatti, M., and S. Pimpinelli. 1992. Functional elements in *Drosophila melanogaster* heterochromatin. Annual Review of Genetics **26:** 239-275.

Hughes J. F., H. Skaletsky, T. Pyntikova, T.A. Graves, S. K. van Daalen *et al.* 2010 Chimpanzee and human *Y* chromosomes are remarkably divergent in structure and gene content. Nature **7280:** 536-539.

Huttunen, S. and J. Aspi. 2003 Complex inheritance of male courtship song characters in Drosophila virilis. Behavior Genetics **33:** 17-24.

Karafet, T., P. de Knijff, E. Wood, J. Ragland, A. G. Clark, *et al*. 1998 Different patterns of variation at the *X*- and *Y*-chromosome-linked microsatellite loci DXYS156X and DXYS156Y in human populations. Human Biology **70:** 979-92.

Koerich, L. B., X. Wang, A. G. Clark, A. B. Carvalho, 2008 Low conservation of gene content in the Drosophila *Y* chromosome. Nature **7224:** 949-51.

Krsticevic, F.J., H.L. Santos., S. Januario, C.G. Schrago, and A.B. Carvalho. 2010 Functional copies of the *Mst77F* gene on the *Y* chromosome of *Drosophila melanogaster*. Genetics **184**: 295-307.

Kuijper, B., A. D. Stewart, and W. R Rice. 2006 The cost of mating rises nonlinearly with copulation frequency in a laboratory population of *Drosophila melanogaster*. Journal of Evolutionary Biology **19:** 1795-1802.

Lemos, B., L. O. Araripe, and D. L. Hartl. 2008 Polymorphic *Y* chromosomes harbor cryptic variation with manifold functional consequences. Science **319:** 91-93.

Lloyd, V. K., D. A. Sinclair, and T. A. Grigliatti. 1997 Competition between different variegating rearrangements for limited heterochromatic factors in *Drosophila melanogaster*. Genetics **145:** 945-959.

Lyckegaard, E. M. S., and A. G. Clark. 1991 Evolution of ribosomal-RNA gene copy number on the sex-chromosomes of *Drosophila melanogaster*. Molecular Biology and Evolution **8:** 458-474.

Lyckegaard, E. M. S., and A. G. Clark. 1989 Ribosomal DNA and stellate gene copy number variation on the *Y*-chromosome of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **86:** 1944-1948.

McGraw, L. A., G. Gibson, A. G. Clark, and M. F. Wolfner. 2004 Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*. Current Biology **14:** 1509-1514.

Montchamp-Moreau, C., V. Ginhoux, and A. Atlan. 2001. The *Y* chromosomes of Drosophila simulans are highly polymorphic for their ability to suppress sex-ratio drive. Evolution **55:** 728-737.

Muller, H. J. 1930. Types of visible variations induced by *X*-rays in Drosophila. Journal of Genetics **22:** 299-335.

Pal-Bhadra, M., B. A. Leibovitch, S. G. Gandhi, M. Rao, U. Bhadra *et al*. 2004 Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science **303:** 669-672.

Repping, S., H. Skaletsky, L. Brown, S. K. M. van Daalen, C. M. Korver, *et al*. 2003 Polymorphism for a 1.6-Mb deletion of the human *Y* chromosome persists through balance between recurrent mutation and haploid selection. Nature Genetics **35:** 247-251.

Rice, W. R. 1987 Genetic hitchhiking and the evolution of reduced genetic activity of the *Y* sex chromosome. Genetics **116:** 161-167.

Robertson, H. M., C. G. Warr, and J. R. Carlson. 2003 Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **100:** 14537-14542.

Rohmer, C., J. R. David, B. Moreteau, and D. Joly. 2004 Heat induced male sterility in Drosophila melanogaster: adaptive genetic variations among geographic populations and role of the *Y* chromosome. Journal of Experimental Biology **207:** 2735-2743.

Rozen, S., J. D. Marzalek, R. K. Alagappan, H. Skaletsky, and D. C. Page. 2009 Remarkably little variation in proteins encoded by the *Y* chromosome's single-copy genes, implying effective purifying selection. American Journal of Human Genetics **85:** 923-928.

Schulze, S. R., and L. L. Wallrath. 2007 Gene regulation by chromatin structure: paradigms established in *Drosophila melanogaster*. Annual Review of Entomology **52:** 171-192.

Shen, P. D., F. Wang, P. A. Underhill, C. Franco, W. H. Yang, *et al*. 2000 Population genetic implications from sequence variation in four *Y* chromosome genes. Proc. Natl. Acad. Sci. USA **97:** 7354-7359.

Spach, K. M., M. Blake, J. Y. Bunn, B. McElvany, R. Noubade, et al. 2009 Cutting Edge: The *Y* Chromosome Controls the Age-Dependent Experimental Allergic Encephalomyelitis Sexual Dimorphism in SJL/J Mice. Journal Of Immunology **182:** 1789-1793.

Spofford, J.B. 1976 Position-effect variegation in Drosophila pp. 955-1018 in *The Genetics and Biology of Drosophila*, Vol 1c, edited by M. Ashburner and E. Novitiski, Academic Press, New York.

Stoltenberg, S. F., and J. Hirsch. 1997. Y-chromosome effects on Drosophila geotaxis interact with genetic or cytoplasmic background. Animal Behaviour **53:**: 853-64.

Storey, J. D. and R. Tibshirani. 2003 Statistical significance of genomewide studies. Proc. Natl. Acad. Sci. USA **100:** 9440-9445.

Talbert, P. B., and S. Henikoff. 2006. Spreading of silent chromatin: inaction at a distance. Nature Reviews Genetics **7:** 793-803.

Teuscher, C, R. Noubade, K. Spach, B. McElvany, J. Y. Bunn, *et al*. 2006 Evidence that the *Y* chromosome influences autoimmune disease in male and female mice. Proc. Natl. Acad. Sci. USA **103:** 8024-8029.

Townsend, J. P., and D. L. Hartl. 2002 Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments. Genome Biology **3:** R0071.

Vibranovski, M. D., L. B. Koerich, and A. B. Carvalho. 2008 Two new *Y*-linked genes in *Drosophila melanogaster*. Genetics 1**79:** 2325-2327.

Voelker, R. A., and K. Kojima. 1971 Fertility and fitness of XO males in Drosophila. I. Qualitative study. Evolution **25:** 119-128.

Volpe, T. A., C. Kidner, I. M. Hall, G. Teng, S. I. S. Grewal, *et al*. 2002 Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science **297:** 1833-1837.

Wallrath, L. L. 1998 Unfolding the mysteries of heterochromatin. Current Opinion in Genetics & Development **8:** 147-153.

Wang, P., R. F. Lyman, T. F. C. Mackay, and R. R. H. Anholt. 2009 Natural variation in odorant recognition among odorant binding proteins in *Drosophila melanogaster*. Genetics **184:** 759-767.

Wu H., M. K. Kerr, X. Q. Cui, and G. A. Churchill. 2003 MAANOVA: a software package for the analysis of spotted cDNA microarray experiments, pp. 313– 431 in *The Analysis of Gene Expression Data: an Overview of Methods and Software*, edited by G. Parmgiani, E. S. Garrett, R. A. Irizarry, and S. L. Zeger. Spring, New York.

Zhang, P. and R. L. Stankiewicz. 1998 *Y*-linked male sterile mutations induced by *P* element in *Drosophila melanogaster*. Genetics **150:** 735-744.

Zurovcova, M. and W. F. Eanes. 1999. Lack of nucleotide polymorphism in the *Y*-linked sperm flagellar dynein gene Dhc-Yh3 of *Drosophila melanogaster* and *D. simulans*." Genetics **153:** 1709-1715.

Chapter 2

Imprinting of the *Y* chromosome provides insight into chromatin-modulating mechanism underlying *Y*-linked regulatory variation in *Drosophila melanogaster*

Pan-Pan Jiang and Daniel L. Hartl

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge MA 02138

ABSTRACT

Our study describes a novel report of *Y* chromosome parent-of-origin imprint effects on global gene expression in *Drosophila melanogaster*. This supports past empirical work indicating that imprinting effects in this species are limited to heterochromatic regions of the genome, in particular, on the *Y* chromosome, which constitutes a large heterochromatic body. First, we show that maternal imprinting of the *Y* chromosome reduces the extent of *Y*-linked regulatory polymorphism. Second, rDNA transcription levels are correlated with a heterochromatin-related phenotype (position effect variegation). And third, there is a polymorphic imprint effect such that we can detect a significant *Y*-by-parent interaction effect when comparing two different Y chromosomes inherited either maternally or paternally. The genes showing this interaction effect are associated with essential male functions including spermatogenesis and specific expression in testes. This, together with previous research suggesting that imprinting is linked to changes in chromatin states, supports the hypothesis that the *Y* chromosome is acting as an important determinant of global gene expression diversity in males via chromatin remodeling mechanisms. We believe that naturally occurring *Y* chromosomes are polymorphic for elements that can regulate the expression of biologically relevant phenotypes by modulating a process that is changed by parent-of-origin imprinting, namely, the manipulation of global chromatin dynamics.

INTRODUCTION

The *Y* chromosome is a genetic material found only in males and represents a unique genomic opportunity for the accumulation of genes benefiting males and the resolution of genetic conflicts (Rice et al. 2008). However, theoretical expectations of a broad role for the *Y* have not been matched by empirical findings due to the unusual molecular characteristics of this chromosome (with a few notable exceptions (Carvalho et al. 1997; Rice 1998; Chippindale and Rice 2001)). Because the *Y* chromosome is transmitted without recombination from father to son, complete linkage between genes results in the accumulation of deleterious alleles and the loss of genetic diversity (Bull 1985; Rice 1987; Charlesworth and Charlesworth 2000; Bachtrog et al. 2008). In *Drosophila melanogaster*, the *Y* chromosome accounts for ~40 Mb of DNA (>20% of the haploid genome) (Carvalho 2002; Carvalho et al. 2009), but of the approximately 14,000 genes in the *D. melanogaster* genome, only 13 (~0.1%) have been mapped to the *Y*, some of them encoding proteins for spermatogenesis-specific processes and others with functions yet unknown (Carvalho et al. 2009; Krsticevic et al. 2010). Within these few *Y*-linked protein-coding genes, nucleotide diversity is effectively zero (Zurovcova and Eanes 1999), leading to the widespread belief that *Y*-linked genes are monomorphic within the species. This pattern of low gene density and nucleotide diversity on the *Y* holds in most animal species, including humans, where the *Y* chromosome harbors very few protein-coding genes in the male-specific region of the *Y*, and these genes also have very low polymorphism (Shen et al. 2000; Rozen et al. 2009).

Despite the lack of sequence polymorphism in *Y*-linked genes, many *Y*-linked phenotypic traits in *Drosophila* have been observed. They include thermal sensitivity of

spermatogenesis (Chakir et al. 2002; David et al. 2005), male fitness (Chippindale and Rice 2001), reproductive isolation (Bayes and Malik 2009), sex ratio distortion (Carvalho et al. 1997; Montchamp-Moreau et al. 2001), sexually antagonistic zygotic drive (Rice et al. 2008), heterochromatic silencing (Dimitri and Pisano 1989; Dorer and Henikoff 1994), genetic imprinting (Maggert and Golic 2002; Menon and Meller 2009), and courtship behavior (Huttunen and Aspi 2003). In addition, *Y*-linked interpopulation polymorphisms influence the expression of many genes located on other chromosomes, many of which are connected to male fitness (Sackton et al. 2011; Lemos et al.). In light of this seeming abundance of phenotypic variation associated with *Y*-polymorphism, the lack of nucleotide diversity in *Y*-linked genes seems incongruous. However, looking at structural variation on the *Y*, instead of sequence variation, may reconcile this conundrum.

Structurally, the *Y* chromosome is unique because it is predominantly heterochromatic and highly enriched in repetitive sequences (Weiler and Wakimoto 1995). Heterochromatin constitutes 20-30% of human and fly genomes (Smith 2007), and up to 85% of other genomes (Moritz and Roth 1976), but heterochromatin is still relatively poorly understood. The classic understanding of heterochromatin as transcriptionally repressive and inert is overly simplistic, as genes that natively reside within euchromatin can be silenced when transposed to heterochromatin, and vice versa (Hearn et al. 1991). Unlike the lack of diversity found in *Y*-linked, protein-coding genes, *Y* chromosome structural differences are pronounced: heterochromatic and rDNA repeats in Drosophila can differ in repeat number or length between different *Y* lineages (Lyckegaard and Clark 1989). Manipulating the amount of *Y*-linked heterochromatin can

affect heterochromatin-related phenotypes elsewhere. The most commonly studied heterochromatin-associated phenotype is position effect variegation (PEV), in which a classic example is the white-eye gene being repositioned from its native position in euchromatin near the tip of the *X* chromosome to a new position close to the AT-rich microsatellite region in the pericentromeric heterochromatin of the *X* (Muller 1930). This repositioning subjects the white-eye gene to the effects of heterochromatin spreading, and results in an eye color phenotype that is a mosaic of red and white. Manipulations that increase the amount of *Y*-linked heterochromatin lead to decreased heterochromatinization near the euchromatin-heterochromatin boundary where the white-eye gene is repositioned (Dimitri and Pisano 1989), which suggests the redistributional effect of a fairly constant supply of chromatin regulators. This observation also supports the locking-molecule model proposed by Zuckerkandle (Zuckerkandl 1974) that certain chromatin molecules are present in limited amounts, and act via repetitive sequences to cause transitions from euchromatic to heterochromatic states. This model was proposed in opposition to the simplistic view of chromosomes as being deterministically divided into two defined states. In further support of the idea that *Y*-heterochromatin is involved in *Y*-linked phenotypic variation, polymorphic *Y*'s sampled from geographically diverse *D. melanogaster* populations have differential effects on PEV (Lemos et al. 2008; Zhou et al. 2012) and enriched effects on genes located near the *X* chromosome euchromatin-heterochromatin boundary (Jiang et al. 2010).

Heterochromatin formation also affects rDNA arrays (Peng and Karpen 2007), with partial deletions of rDNA arrays resulting in reduced heterochromatin-induced gene silencing (Paredes and Maggert 2009). Overall heterochromatin levels in a cell are

difficult to quantify, however, the link between rDNA arrays and heterochromatin may provide an estimate: a recent study found that rRNA transcription indicates reduced heterochromatin in the genome (Larson et al. 2012). Taken together, these results suggest that polymorphisms in rDNA copy number might be relevant to global chromatin structure and concomitant gene regulation, and explains findings that *Y*-linked rDNA variation is an important determinant of global gene expression diversity in *D. melanogaster* (Paredes et al. 2011).

Both males and female fruit flies can establish and maintain epigenetic chromatin states, and cases are known in other insects where males and females establish different epigenetic states in meiosis, termed imprints, resulting in parent-of-origin-dependent chromatin behavior. Imprinting in *Drosophila* has been mainly detected through modulation of PEV (Golic et al. 1998; Haller and Woodruff 2000; Maggert and Golic 2002; Menon and Meller 2009). Heterochromatin formation is required for maintenance of imprint marks in zygotes and plays a central role in imprinting (Lloyd 2000). Because the *Y* chromosome constitutes a large heterochromatic body full of repetitive sequences, it has been proposed to act as a sink for chromatin modulators and other DNA-binding proteins which use repeat sequences as binding motifs, such as the GAGA transcription factor (Weiler and Wakimoto 1995; Lemos et al. 2008), which may effect establishment of chromatin states.

Sex chromosomes are often studied for imprints because their inheritance is predictable. In female marsupials and in the placental tissues of female rodents, where dosage is compensated by *X*-inactivation in females, the paternal *X* is preferentially inactivated during dosage compensation (Migeon 1998). In *Drosophila,* this type of

specific inactivation is not expected since both *X* chromosomes are expressed in females, and dosage is compensated in males by overexpression of the *X*. In this genus, imprinting does not identify the *X* chromosome, as males are able to dosage compensate normally regardless of the origin of their *X* chromosome (Menon and Meller 2009). Few examples of autosomal imprints have been found, as experiments with deletions throughout the genome found uniparental diploids of *D. melanogaster* to be viable and to have no visible phenotype (Lindsley and Grell 1969). However, the *Y* chromosome presents a different story, as evidence from marker gene and PEV studies seem to suggest that the *D. melanogaster Y* is under imprint-regulated control, unlike the *X* or autosomes (Golic et al. 1998; Haller and Woodruff 2000; Maggert and Golic 2002). Reversal of sex chromosome inheritance can reveal imprinting effects and has helped clarify the role of the *Y* in dosage compensation, where it likely modulates *X* chromosome recognition or the ability of the *male-specific lethal* complex to modify chromatin (Menon and Meller 2009). Previous work on imprinting in *Drosophila* has identified the imprint to reside only within heterochromatic, gene-poor regions, and thus, "while imprinting occurs in *Drosophila*, it is generally without phenotypic consequence" (Lloyd 2000). However, in light of our new understanding of the important role of the *Y* chromosome in regulating non-*Y* gene expression, and its potential to facilitate male-specific evolution, the imprinting of large heterochromatic bodies may have significant evolutionary consequences. We hypothesized that reversal of sex chromosome inheritance, if it disrupts the normal *Y* chromosome imprinting, may also lead to differential *Y*-linked regulatory variation between polymorphic *Y*'s.

In accordance with our hypothesis, we observed that reversed sex chromosome inheritance can dramatically alter the scope and magnitude of *Y*-linked regulatory variation in *D. melanogaster*. In agreement with previous findings, males carrying a *Y* chromosome derived from a population collected in the Democratic Republic of the Congo, Africa (*Ycon*) or a *Y* chromosome from a laboratory population collected in Ohio, USA (*Ycs*) showed differential expression, as determined by microarray analysis, at hundreds of autosomal and *X*-linked loci. We then exploited the fact that in diploid *Drosophila,* sex is determined not by the *Y*, but by the number of *X* chromosomes present (Baker and Belote 1983), and males can inherit their *Y*'s from *XXY* mothers (carrying an attached-*X* and a *Y* chromosome), and an *X* from normal *XY* males. These males also carried either *Ycon* or *Ycs*, but experienced reversed sex chromosome inheritance. Differential gene expression between these males was much reduced. Most interestingly, we found a set of hundreds of genes showing differential *Y*-by-origin interaction effects, with enrichment in this set for genes related to ribosomal function and male-specific traits. In addition, the effect of *Y*-by-origin interaction effects on PEV and rDNA transcription levels, two traits with demonstrated links to heterochromatin, were strongly correlated and consistent. We propose that parent-of-origin imprints established on the *Y* likely influenced the establishment of chromatin states on the chromosome, and altered the titration of available chromatin proteins within the nucleus, thereby influencing gene expression at multiple other loci whose expression is regulated, at least partially, at the structural, chromatin level.

METHODS AND MATERIALS

Drosophila stocks. Two strains of diverse origin were used as donors of the *Y*-chromosome: the common laboratory strain Canton-S (collected in Ohio in the 1940's) and Tucson Drosophila Stock Center strain number 14021 (collected in the Democratic Republic of the Congo). Males from these strains were used for their *Y* chromosomes to be introgressed into the same laboratory stock background (BL4361, previously described) (Lemos et al. 2008; Jiang et al. 2010; Lemos et al. 2010). These crosses resulted in two isogenic cultures from which we could sample the Y chromosome, either *Ycs* or *Ycon*. Since Y chromosomes in these crosses were inherited normally, from the paternal lineage, we call them *Ycs*-pat and *Ycon*-pat. Crosses to obtain stocks with reverse sex chromosome inheritance resulting in males with identical genetics as *Ycs* and *Ycon*, but with a maternally inherited *Y* and paternally inherited *X*, are shown in Fig. S1. These crosses provided us with two other isogenic cultures from which we could sample *Y* chromosomes with sex-reversed inheritance, either *Ycs*-mat or *Ycon*-mat. Flies were grown under 12 hour light/dark temperature and humidity controlled incubators at 25 ºC. Males from the isogenic stocks were collected for use in microarray dye-swap experiments. Newly emerged males were collected 1-3 days post eclosion and allowed to age for 2 days at 25C, after which they were flash frozen in liquid nitrogen and stored at -80C.

Microarray hybridization and analysis. Our experiment consisted of 12 cDNA microarrays, involving 24 separate labeling reactions. Array design (Figure S2.2) was selected to optimize detecting interaction effects of *Y* chromosome and sex chromosome inheritance pattern (paternal-derived or maternal-derived *Y*). Microarrays were ~18,000-

feature cDNA arrays spotted with *D. melanogaster* cDNA PCR products as described (Lemos et al. 2008). RNA extraction, cDNA synthesis, microarray hybridization, and microarray slide scanning protocols closely followed that of Lemos et al (Lemos et al. 2008). Stringent quality-control criteria were used to ensure reliability of foreground intensity reads for both Cy5 and Cy3 channels. Foreground fluorescence of dye intensities was normalized by the Loess method implemented in the library Limma of the statistical software R. In cases where multiple probes map to a single FBgn, we selected a single probe for each FBgn by first selecting the probe that had good quality data on the most arrays, and then selecting the probe with the highest average expression level. In total, we included data from 5652 probes. Microarray gene expression data herein reported can be obtained at the GEO database (GSE42157).

We fit a linear model using Limma method lmfit (Smyth 2004), with a design matrix parameterized to include a Dye term. To detect Y chromosome by parent of origin (*Y*-by-origin) interaction effects, we fit the contrast (*Ycon*-mat – *Ycs*-mat) – (*Ycon*-pat – *Ycs*-pat*)*. Positive fold-change values represent higher relative expression in *Ycon* compared to *Ycs* males when sex chromosome inheritance is reversed (maternal-derived *Y*), than when sex chromosome inheritance is normal (paternal-derived *Y*). We call these Type 1 interaction effects. Negative fold-change values represent the reverse trend. We call these Type 2 interaction effects. All *P* values from Limma are adjusted for multiple testing using the false discovery rate (FDR) approach (Benjamini and Hochberg 1995). We focus on the 20% FDR dataset, because it aligns with similar FDR cut-offs used by us in the past, but our conclusions are not substantially affected by using a different cutoff.

Gene expression datasets. Tissue-specific gene expression patterns were downloaded from FlyAtlas, and the tissue specificity of each gene was ascribed as outlined in (Sackton et al. 2011). The measure of tissue specificity index ranges from 0 if a gene is not expressed in a tissue, to 1 if a gene is exclusively expressed in a tissue. A gene is defined as being specific to a tissue when the tissue specificity is ≥ 0.9.

Male and female-biased gene expression data were downloaded from SEBIDA (Gnad and Parsch 2006). In the SEBIDA meta-analysis of relative expression level of genes in males and females, a 2-fold cut-off and FDR > 0.2 was used to assign genes to categories of male-biased or female-biased. Male-fitness, female-fitness, and sexually-antagonistic fitness genes were those identified by Innocenti and Morrow (Innocenti and Morrow 2010) to be associated with fertilization success in males and females, under competitive conditions.

GO Enrichment. Enriched GO categories were identified with the term-enrichment tool available as part of AmiGO (http://amigo.geneontology.org), using a Bonferroni-corrected *P*-value cutoff of 0.05, and using the set of 5652 probes passing our quality control filters as background set.

Quantitative-PCR. We used the non-LTR retrotransposon element *R2* to determine the level of rDNA transcription. *R2* retrotransposons insert specifically into the 28S rRNA gene and are cotranscribed with the rRNA units, and thus can be used as a sensitive marker for rDNA transcription. *R2* was amplified with primers 5'-

TTGAGAGCAGAGGGGGAGTA-3' and 3'- GTTTAGCATTACCGGGACCA-5'. For quantitative analysis, three to four biological replicates of 60 adult flies each were sampled in each genotype. qPCR analyses were carried out with Fast Sybr Green Master Mix (Applied Biosystems). cDNA synthesis was done with Quantitect Reverse-Transcription Kit (Qiagen). Real-time PCR profiles were obtained with 7900HT Fast Real-Time PCR (Applied Biosystems). *Beta-tubulin* expression level was used as control.

PEV assay. Two sets of males were produced. 1) Males with *Ycs* or *Ycon* were crossed to females from a stock carrying the *white* eye-color gene *w[m4h]* to produce male progeny variegating for *white*, and with a paternally inherited *Y.* 2) Females (with compound *X* chromosome and *Y*) from the populations containing *Ycs*-mat and *Ycon*-mat were crossed to males from the same *w[m4h]* stock to produce male progeny also variegating for *white,* but with a maternally inherited *Y.* Cross designs are shown in Figure S2.3. Flies from this strain possess an inversion on the *X* chromosome that repositions the *white* gene proximal to the *X* centromere. Due to its new location near the heterochromatin-euchromatin boundary, *w[m4h]* experiences variegated gene silencing, resulting in eye-color mosaicism of mottled white and red eye spots. Flies were maintained at 25 ºC. Males from these crosses were collected, aged for 2 days at 25 ºC, flash frozen in liquid nitrogen, and stored at -80 ºC. Heads of males were removed with a blade and homogenized 4 to a tube of 10 uL of acidified ethanol (30% ethanol acidified to pH 2 with HCl). Eye pigment expression was assessed with spectrophotometric analysis at an optical density of 280 nm. Four to six biological replicates were used per treatment, with two technical replicates per biological replicate. Males displaying typical eye-

pigmentation phenotypes were imaged using an automontage system (Syncroscopy, Frederick, MD).

To test whether maternal aneuploidism, rather than maternal transmission of the Y chromosome, was a major contributor to PEV effects, we performed crosses outlined in Supplementary Figure S2.4. Briefly, males inherited identical Y chromosomes either maternally from attached-XX/Y mothers, or paternally from XY males (as normal), but also had genetically XXY mothers (in this case they were attached-XY/X). If maternal versus paternal transmission of the Y chromosome was a major contributor to differential PEV expression, then we would expect to see differences in PEV for these males. If maternal aneuploidism (XXY) versus normal XX maternal effects were the cause of major PEV differences, we would not see differential PEV phenotypes since both males had XXY mothers.

RESULTS

***Y*-regulatory variation is modulated by parent -of-origin of the *Y* chromosome**. Previous work suggested that hundreds of genes are under the regulatory control of the *Y* chromosome in *Drosophila melanogaster* (Lemos et al. 2008; Jiang et al. 2010; Sackton et al. 2011), but the mechanism of such regulation remained elusive. Here, we show that males inheriting a *Y* chromosome from their father, and an *X* from their mother, showed differential YRV in both magnitude and scope of genes affected. At false discovery rates FDR < 0.2, males possessing either *Ycon* or *Ycs* through normal sex chromosome inheritance (*Ycon*-pat vs *Ycs*-pat) showed differential expression of 242 genes, while the

difference between males with *Ycon* or *Ycs* acquired through maternal inheritance (*Ycon*-mat vs *Ycs*-mat) showed differential expression of only 11 genes. The number of genes showing significant *Y*-by-origin interaction, representing genes which show significantly increased, decreased, or reversed magnitudes of differential expression between *Ycs* and *Ycon* males depending on paternal or maternal inheritance patterns, totaled 392 genes. Overlap of differentially expressed gene sets between these three contrasts totals 3 genes, as shown in Figure 2.1. These results suggest that epigenetic modifications established during transmission of sex chromosomes differs between males and females, with important implications for YRV.

Our observation of reduced *Y*-linked regulatory variation between males carrying polymorphic *Y* chromosomes inherited from their mothers may be explained by an overall paternal imprint on the *Y* in *Drosophila*, which leads to increased heterochromatinization of the chromosome (Golic et al. 1998; Maggert and Golic 2002). When *Y* chromosomes are maternally inherited instead, heterochromatinization may be reduced. In light of the hypothesis that chromatin-based regulation by the *Y* is mediated by repeated sequences acting as a sink for chromatin-associated proteins, maternally inherited *Y* chromosomes may not modulate chromatin dynamics as pervasively as paternally inherited *Y* chromosomes. The limited role of maternally inherited *Y* chromosomes as modifiers of global gene expression is not surprising, given that *Y* chromosomes in *Drosophila* normally never pass through the female lineage.

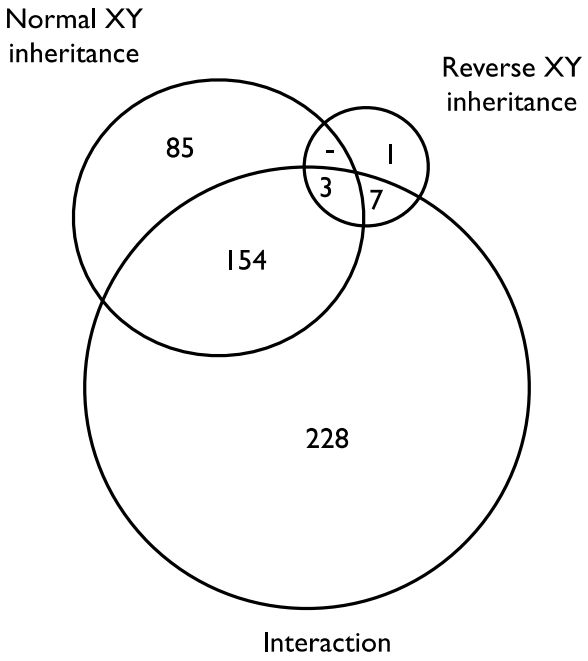


Figure 2.1. Venn diagram showing number of differentially regulated genes between *Ycs* and *Ycon* males with sex chromosomes inherited normally (X-mat, Y-pat) or reverse (X-pat, Y-mat) (left and right top circles). Lower circle represents genes which show significant *Y*-by-origin interaction effects.

We propose that the 392 genes showing *Y*-by-origin interaction effects represent genes which show differential Y-regulatory variation because of an imprint on the Y caused by its parent-of-origin. They can be subdivided into two categories (outlined in Methods), depending on their expression patterns: 1) Type 1 interaction effects represent *higher* relative expression in *Ycon* compared to *Ycs* males when sex chromosome inheritance is reversed (maternal-derived *Y*), than when sex chromosome inheritance is normal (paternal-derived *Y*), 2) Type 2 interaction effects *lower* relative expression for the same contrast. There were 247 genes showing Type 1 effects, and 145 genes showing Type 2 effects. A heat map of the relative expressions of these genes across the four *Y*-by-origin combinations is presented in Figure 2.2, with a heuristic representation of "Type 1" vs "Type 2" interaction effects shown for illustrative purposes. Figure 2.3 shows expression patterns of the three genes identified to show differential gene expression in all three contrasts (three-way overlap in Figure 2.1 Venn diagram). These three genes are *CG1979,* a gene with unknown molecular function that has high expression in male testis; a gene for ribosomal protein S10b, a constituent of ribosomes; and a gene for adult cuticle protein 65Aa, a constituent of chitin-based cuticle. Approximately 45%, 42%, 12%, and <1% of the genes showing *Y*-by-origin interaction effects can be assigned to genes on the second, third, *X*, and fourth chromosomes, respectively. These percentages do not differ significantly from expectations based solely on gene content (chi-square test, chi-square = 0.39, *P* value = 0.94).

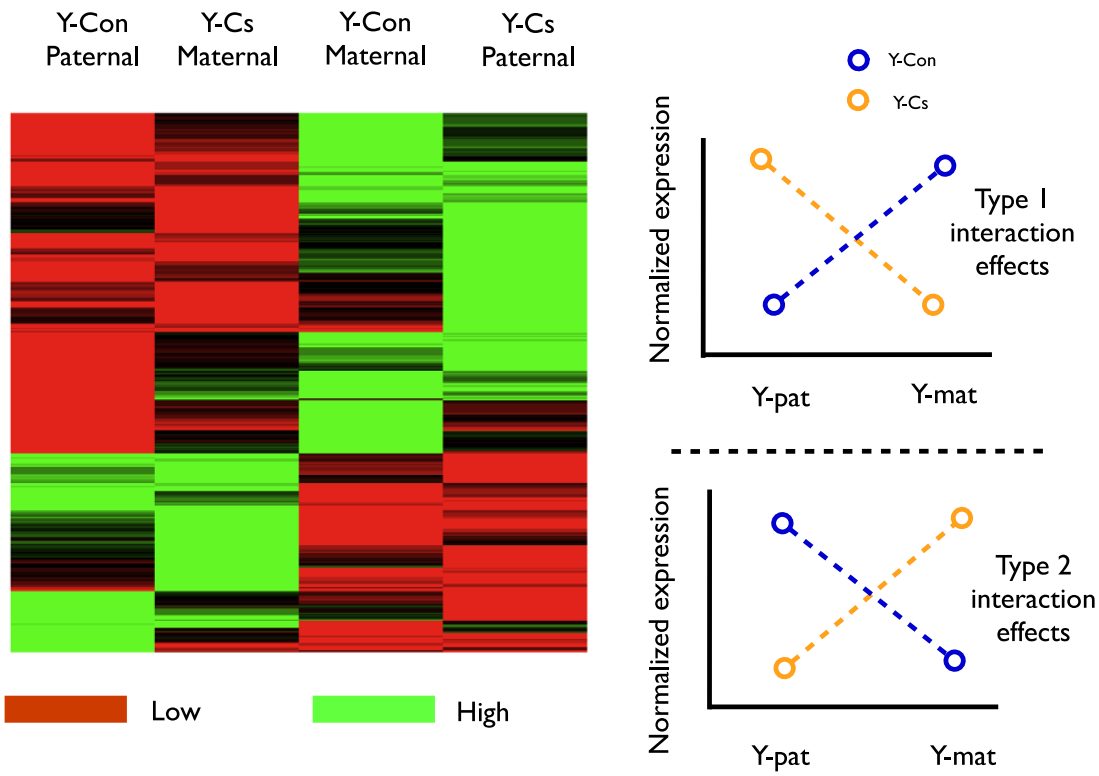


Figure 2.2. Left: A heat map of relative expression levels of 392 genes showing significant *Y*-by-origin interaction effects. Red denotes low expression while green denotes high. Genotypes are given for each column at the top. Rows represent genes. Genes above the dashed line (right panel) demonstrate what we term a "Type 1" interaction effect, while genes below the dashed line demonstrate a "Type 2" interaction effect. Right: Cartoon gene expression patterns of two example genes are given for a heuristic illustration of what these labels for Type 1 and Type 2 interaction gene expression patterns represent. Y-axis shows normalized gene expression expression. X – axis shows the parent of origin of the Y, which is inherited either paternally (*Y*-pat) or maternally (*Y*-mat)

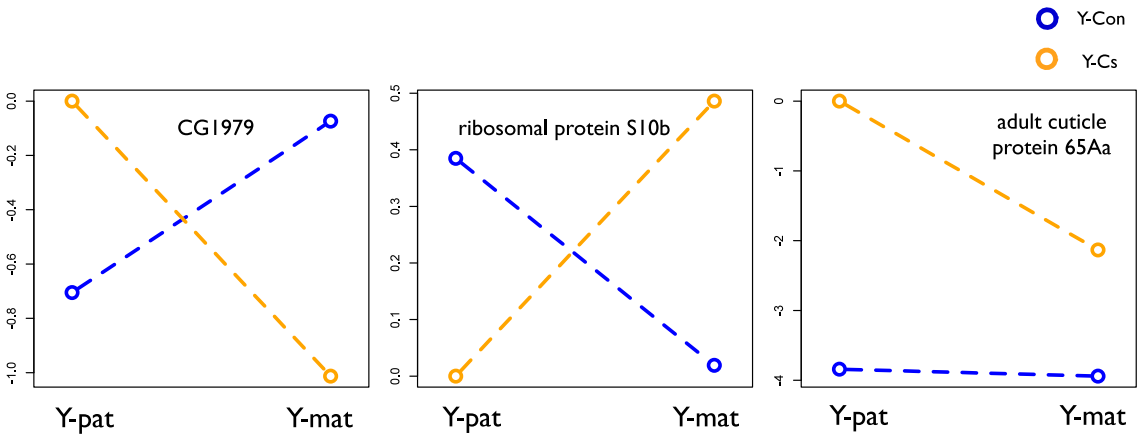


Figure 2.3. Normalized expression levels of *Ycon* and *Ycs* males inheriting their sex chromosome either paternally or maternally. *CG1979* shows Type 1 interaction effects, the gene for ribosomal protein S10b shows Type 2 interaction effects, and the gene for adult cuticle protein 65Aa shows Type 1 interaction effects.

Table 2.1. Significantly over-represented Gene Ontology (GO) categories among genes showing Type 2 *Y*-by-origin interaction effects. P-values are adjusted for multiple hypothesis testing. There are no GO categories significantly enriched among genes showing Type 1 interaction effects.

Category	Description	P-value	Sample frequency	Background frequency
Biological Process				
GO: 0006412	translation	7.47E-04	22/120 (18.3%)	252/4473 (5.6%)
Cellular Component				
GO:0044391	ribosomal subunit	8.3E-07	19/120 (15.8%)	128/4473 (2.9%)
GO:0005840	ribosome	1.43E-06	19/120 (15.8%)	132/4473 (2.9%)
GO:0030529	ribonucleoprotein complex	6.12E-05	24/120 (20.0%)	258/4473 (5.8%)
GO:0015934	large ribosomal subunit	1.27E-04	13/120 (10.8%)	78/4473 (1.7%)
GO:0022626	cytosolic ribosome	4.07E-04	12/120 (10.0%)	72/4473 (1.6%)
GO:0044445	cytosolic part	7.69E-03	12/120 (10.0%)	94/4473 (2.1%)
GO:0022625	cytosolic large ribosomal subunit	2.16E-02	8/120 (6.7%)	44/4473 (1.0%)
GO:0005576	extracellular region	4.11E-02	19/120 (15.8%)	252/4473 (5.6%)
Molecular Function				
GO:0003735	structural constituent of ribosome	1.25E-06	19/120 (15.8%)	131/4473 (2.9%)
GO:0005198	structural molecule activity	1.28E-04	21/120 (17.5%)	209/4473 (4.7%)

Gene expression patterns for genes showing positive *Y*-by-origin interaction effects are involved in ribosome formation and male-specific functions. Gene ontology analysis revealed a functionally cohesive story in regard to which categories were enriched for genes showing *Y*-by-origin interaction effects (Table 1). Genes showing Type 1 interaction effects are functionally distinct from those showing Type 2 interaction effects. Genes in the Type 2 interaction effects group show functional cohesiveness, and are enriched in GO categories related to ribosome function or whose products localized to ribosomal subunits. This pattern of enrichment of rDNA-based functional units again underlies the importance of heterochromatin and chromatin structure to *Y* chromosome regulation. rDNA transcription is necessary for heterochromatin formation, and indicative of overall heterochromatinization of the genome (Peng and Karpen 2007; Larson et al. 2012).

Although no GO categories are significantly over-represented among genes in the Type 1 interaction effects group, we do find other interesting patterns of gene expression. For example, genes showing positive interaction effects are more likely to be testis specific than genes showing no or negative interaction effects (Figure 2.4). This is in agreement with previous findings (Lemos et al. 2008; Sackton et al. 2011) of strong YRV effects on male-specific genes and genes associated with spermatogenesis. Based on SEBIDA's meta-analysis on genes which show sex-biased expression (Gnad and Parsch 2006), genes showing Type 1 interaction effects were more likely to be enriched for genes that are male biased, and conversely, diminished for genes that are female biased (Figure 2.5a; Chi-sq test, $\chi 2 = 25.18$, P value = 3.4e-06). Genes showing Type 2 interaction effects did not differ from neutral expectations (Figure 2.5b; chi-square test,

$\chi 2 = 0.026$, P value = 0.99). In concordance with this finding that male-specific traits are affected, we found that genes known to be associated with male fitness (Innocenti and Morrow 2010) are overrepresented among genes identified to have the largest interaction effects (mean-rank gene set enrichment test, $P = 0.031$).

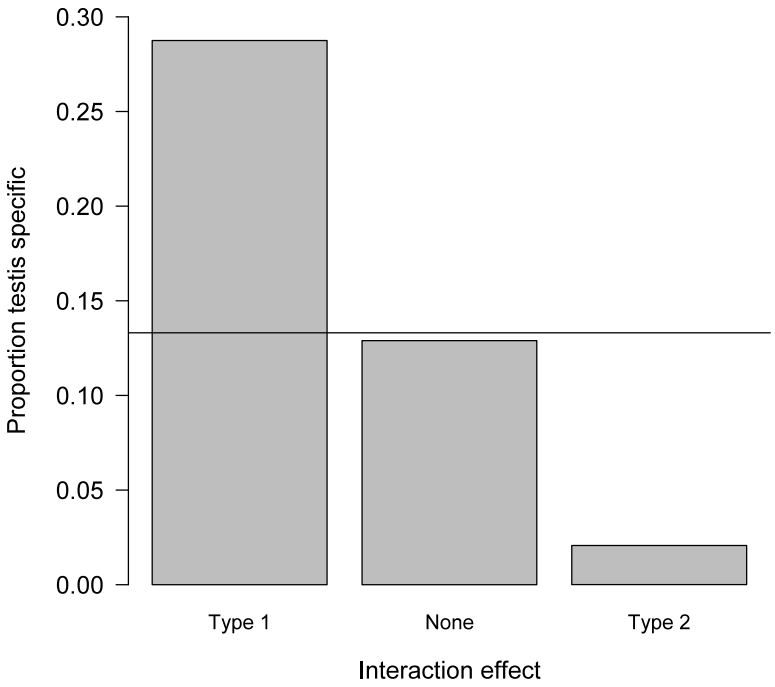
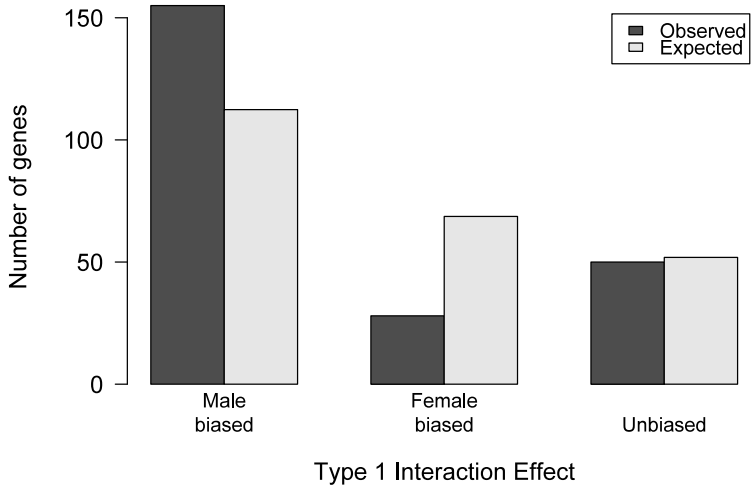


Figure 2.4. Genes showing Type 1 *Y*-by-origin interaction effects are much more likely to be testis specific than genes showing a Type 2 and no interaction effects. Horizontal line represents the expected proportion of testis-specific genes. Fisher's exact test, two-tailed, $P = 1.46e\text{-}06$.

Figure 2.5. (A) Genes showing a Type 1 *Y*-by-origin interaction effect are enriched for male biased genes, but impoverished for female biased genes. (B) Genes showing a Type 2 interaction effect show no deviation from expectation for male biased, female biased, and unbiased genes.

A



B

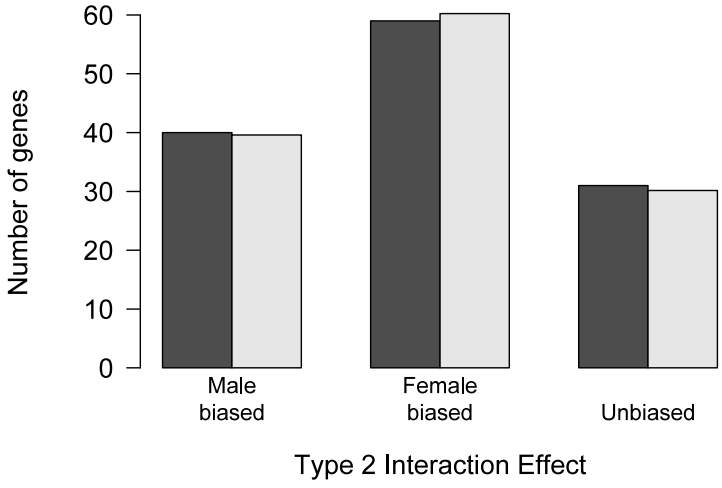
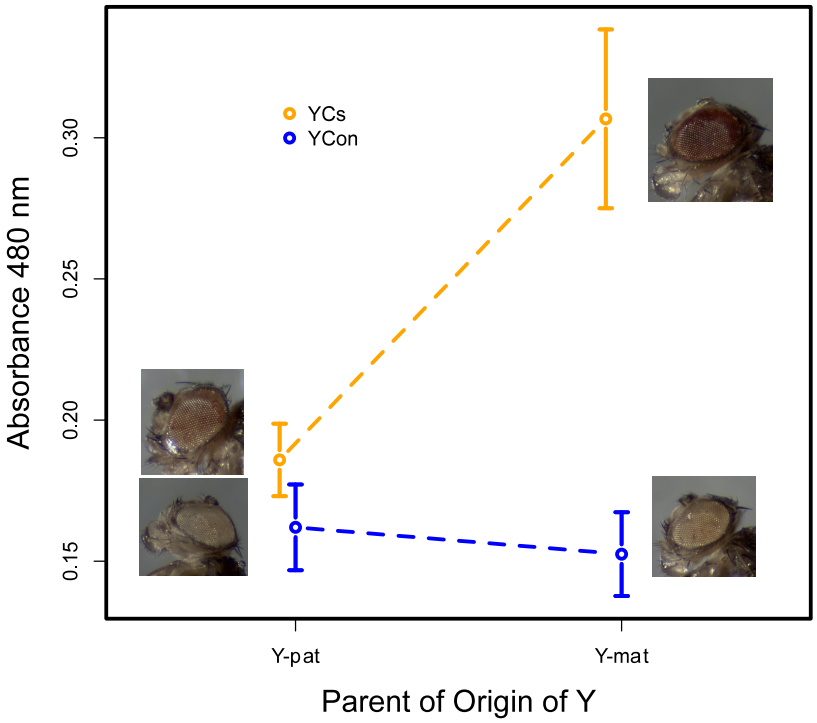


Figure 2.5 (continued).

Parental origin of sex chromosomes affects *Y*-regulated position effect variegation and rDNA transcription levels. The above results from microarray data suggesting *Y*-by-origin interaction effects on gene expression were also observed in a heterochromatin-related phenotype: position effect variegation (PEV). Males in this assay possessed either *Ycon* or *Ycs* in a hybrid genetic background consisting of white-eye mutation *w[m4h]* positioned near the *X* chromosome euchromatin-heterochromatin boundary, a haploid autosomal genome sampled from the original stock containing the PEV marker, and a haploid autosomal genome sampled from the B4361 laboratory population. An additional factor was inheritance of sex chromosomes, which was either normal (*Xmat, Ypat*) or reversed (*Xpat, Ymat*). Reversed sex chromosome inheritance was achieved by crossing *XXY* females (carrying *Ycs* or *Ycon*) with *XY* males carrying the *white* mutation. PEV has previously been shown to be affected by variation in the *Y* (Jiang et al. 2010; Lemos et al. 2010). We observed that *Y*-regulated PEV differs depending on the parent of origin of the *Y* chromosome (Figure 2.6a; ANOVA, $df = 1$, $F = 9.05$, $P = 0.009$), with *Ycs*-mat males showing a marked reduction of heterochromatic-linked silencing (reduced PEV - thereby increasing expression of *white* and displaying redder eyes) compared to *Ycon*-mat males, when sex chromosome inheritance was reversed. When sex chromosome inheritance is normal, *Ycs*-pat males showed slightly higher expression of *white* than *Ycon*-pat males, in accordance with previous studies (Lemos et al. 2010). The effect of *Y* imprinting on PEV and rDNA transcription levels is dependent on the specific *Y* chromosome, as it decreases PEV (resulting in redder eyes) and increases rDNA transcription in *Ycs* males under maternal imprinting, while it has no, or even the opposite, effect in *Ycon* males.

Figure 2.6. (A) *Y*-by-origin interaction effects on PEV. Eye pigmentation was measured as absorption of light at 480 nm. Heads of representative males are shown. (B) Normalized real-time PCR of retrotransposon *R2* expression, a marker of rDNA transcription, in *Ycs* and *Ycon* males inheriting their sex chromosomes paternally or maternally. *Beta-tubulin* expression levels was used as a control.

A



B

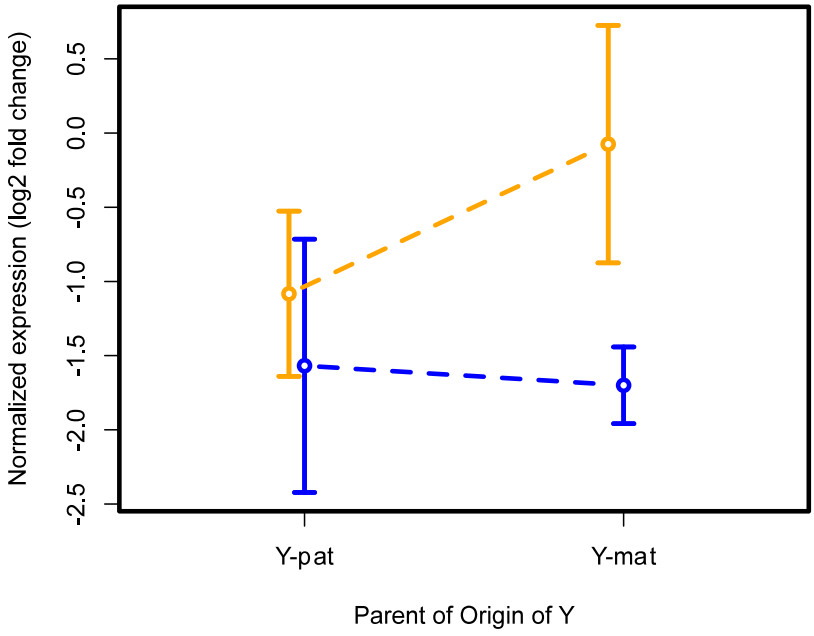


Figure 2.6 (continued).

Interestingly, *Y*-by-origin interaction effects on rDNA transcription levels mirror closely the PEV results, reinforcing the possibility that epigenetic modifications on the *Y* during parent to offspring transfer of genetic material may modify a suite of expression pattern differences. However, the *Y*-by-origin interaction effect on rDNA transcription is not significant (Figure 2.6b; ANOVA, $df = 1$, $F = 0.80$, $P = 0.39$), so the parallelism could be coincidental.

In addition, we observed that males inheriting their Y's maternally had expressed lower levels of PEV than males inheriting their Y's paternally. In contrast to our main PEV study, both these males had genetically XXY mothers (Supplementary Figure S2.5). Thus, paternal versus maternal transmission of the Y chromosome, rather than maternal aneuploidism, is likely the main cause of differential PEV expression.

DISCUSSION

Our results show that imprinting of the *Y* chromosome in *Drosophila melanogaster* affects *Y*-linked regulatory variation in hundreds of genes, as well as position effect variegation (PEV), a heterochromatin-linked phenotype. Convincing evidence of *Y* imprinting in this species has been previously reported. This imprint has usually been detected through expression of genes located on the same chromosome as the imprint, for example, that of the rearranged *Dp(1;f)LJ9* mini-*X* chromosome (Anaka et al. 2009), and *P*-element marker insertions on the *Y* chromosome (Haller and Woodruff 2000; Maggert and Golic 2002). Recently, the imprint effect of the *Y* chromosome has been expanded to include dosage compensation, which is targeted to a chromosome other than the one bearing the imprint (Menon and Meller 2009). Here, we present novel

evidence that imprinting of the *Y* chromosome affects not only proximal genes on the same chromosome, or a dosage effect on the *Y*, but also the expression of a large set of autosomal and *X*-linked genes, many of which have important functional consequences for males.

The biological relevance of the *Y* chromosome imprint, however, remains unclear as *Y* chromosomes are inherited uniparentally in nature. Furthermore, males are fertile whether *Y*'s are transmitted from mothers or fathers, so the imprint is not essential for spermatogenesis or viability.

A link between imprinting and heterochromatin exists in Drosophila. Imprinting in *Drosophila* seems to be associated only with heterochromatin or regions with unusual chromatin structure (Lloyd 2000), such as pericentromeric heterochromatin and the *Y* chromosome. Furthermore, genes implicated in the regulation of imprinting, such as *Su(var)3-9 histone 3 methyl transferase* and *heterochromatin protein 1 (HP1)* are also known to be important regulators of heterochromatin formation (Joanis and Lloyd 2002). Therefore, our finding of an imprint effect on *Y*-regulatory variation further strengthens the hypothesis that *Y*-regulated gene expression on the autosomes and *X* chromosome is exerted by the polymorphic functions of *Y* chromosomes through manipulation of global chromatin dynamics, perhaps by acting as a sink for a limited pool of chromatin remodeling proteins.

Multiple lines of evidence support the hypothesis that repeated sequences serve as a sink for some chromatin-associated proteins, including observations in mice that transcription factor C/EBP alpha binds to satellite repeats (Liu et al. 2007) and also in fruit flies, where enhancer-trap analysis revealed a trans regulator of *lacZ* reporter

expression localizing to functionally redundant and spatially dispersed bands on the *Y* chromosome which coincide with GAGA transcription factor binding motifs (Zhang et al. 2000). Finally, Lemos et al ((Lemos et al. 2010)) found that *XXY* females carrying polymorphic *Y*'s showed differential expression of many genes involved with chromosome organization and chromatin assembly. Since *XXY* females do not transcribe *Y* protein-coding genes, this further supports the hypothesis that *Y*-linked regulatory variation is driven primarily by non-coding, repetitive variation buried in heterochromatin. The *Y*-as-chromatin-sink hypothesis, together with the observation that *Y* chromosomes harbor a plentitude of structural, repeat-based variation, suggests that *Y*-linked polymorphic phenotypic effects may be caused by differential binding of gene regulators by different *Y* chromosomes. This in turn leads to differential titration of these proteins at other genomic loci, and facilitates differential gene expression at these loci. In accordance with this hypothesis, observations of an imprint on *Y* affecting dosage compensation has led to speculation that the imprint could modulate the *Y*'s ability to bind chromatin proteins at a critical time during dosage compensation (Menon and Meller 2009). The link between the *Y* chromosome and chromatin proteins can be extended in humans, where *Y*-chromosome linked repeats are hypomethylated in mutants with compromised global chromatin remodeling (Gibbons et al. 2000).

In general, heterochromatin formed in response to an imprint affects adjacent regions to generate blocks of differential expression according to parent-of-origin (Anaka et al. 2009). The direction of the imprint seems to be mostly paternal imprint, at least for eye and body coloration phenotypes. Studies of the *Dp(1;f)LJ9* mini-*X* chromosome by Anaka et al. (2009) ((Anaka et al. 2009)) found it to be paternally imprinted, with males

exhibiting decreased expression of *y.* Maggert and Golic ((Maggert and Golic 2002)) also found that for 21 out of 22 *Y*-linked *P*-element marker genes they examined, imprint-regulated expression showed the same pattern: paternal inheritance led to reduced expression. We could not test for direct expression of *Y*-linked genes with our experiment because the number of genes is small, and these did not pass quality control filtering. However, results from our PEV assay seem to partially confirm this result: *Ycs*-mat males derived from reverse sex chromosome inheritance crosses (*Y*-maternal, *X*-paternal) showed higher expression of the *white* marker gene. However, *Ycon* males did not show differential expression of *white* between males inheriting their *Y*-chromosomes maternally and those inheriting them paternally. In fact, we would not necessarily expect to see a similar pattern of reduced expression due to paternal-imprint because of higher-level regulatory processes between an imprinted chromosome (in this case, the *Y*), and expression on other chromosomes (in this case, the *X*) in our experiment. In contrast, previous studies had only looked at imprint effects on genes located on the same chromosome as the chromosome of imprint.

Studies of hybrid lethality between *Drosophila* species may provide clues to the adaptive significance of *Y* regulatory variation modulated by *Y*-imprinting. First, hybrid lethality and heterochromatin have been linked in previous studies (Brideau et al. 2006; Pal Bhadra et al. 2006). Second, Menon and Meller (2009) ((Menon and Meller 2009)) found that paternally transmitted *Y* chromosomes dramatically lowers survival in males harboring mutations in *roX1* and r*oX2*, genes necessary for forming the male-specific lethal complex in males responsible for dosage compensation. They proposed that *roX* mutations may mimic a defect in dosage compensation that occurs in interspecies

hybrids, and paternal imprinting of the *Y* is adaptive as it lowers the survival of unfit hybrids (Wu and Ting 2004). Third, in support of this hypothesis, genes under the influence of *Y*-linked polymorphism regulation are highly divergent in expression between species (Lemos et al. 2008), and interspecific *Y* chromosome introgressions led to reduced male fecundity and sperm competitive ability (Sackton et al. 2011). Therefore, a paternal imprint is expected to enhance the regulatory ability of the *Y* chromosome. We find agreement with this hypothesis of a male-adapted paternal imprint on the *Y* within our own data: a particular subset of genes showing *Y*-by-origin interaction effect were enriched for high, specific expression in testis, were more male biased than expected, and were associated with male fitness. In addition, when the *Y* was inherited from mothers, it lost much of its regulatory ability.

Another indication that *Y* chromosome imprinting results in chromatin state remodeling within the cell is the pattern of rRNA synthesis we observed, as quantified by qPCR. Although the interaction effect is not significant, we saw that rRNA transcription levels were modified by reverse sex chromosome inheritance. Previous studies found that rDNA arrays are necessary for heterochromatin formation through two lines of evidence: heterochromatin formation is disrupted in rDNA array mutants (Peng and Karpen 2007) and deletions of the rDNA array results in decreased heterochromatin composition of the genome, and affects gene expression elsewhere in other heterochromatin regions of the genome (Paredes and Maggert 2009). rDNA transcriptional variation is known to be correlated with PEV (Paredes and Maggert 2009; Zhou et al. 2012), presumably because it reflects alteration of chromatin states within the cell, and PEV is a heterochromatin-linked phenotype. We also observe a marked similarity between rRNA expression and

PEV phenotype in our results. In further support of this pattern, Larson et al (2012) ((Larson et al. 2012)) found that increased rRNA transcription indicates a reduced heterochromatic state in the genome. The effect of this on PEV would be to disrupt heterochromatin-induced silencing (reduced variegation), and thus increase expression of *white*, producing redder eyes. In support of our findings of a significant role for rDNA arrays in YRV, Paredes et al (Paredes et al. 2011) recently showed that *Y*-linked rDNA polymorphism modifies the expression of genes in euchromatic components of the *D. melanogaster* genome, and that these genes overlap significantly with genes affected by natural *Y* chromosome polymorphisms.

The potency of *Y* imprint affecting male-specific functions, and regulating global gene expression, is likely not limited to *Drosophila*. The role of the *Y* as a genomic regulatory element has been convincingly demonstrated in murine models. Xu et al ((Xu et al. 2012)) reported that presence of the *Y*, but not absence of the second *X*-chromosome, in *XY* sex-reversed females, alters expression of hundreds of genes. Singer et al ((Singer et al. 2004)) identified the *Y* chromosome to have a phenotypic effect similar to autosomes in a suite of development, physiological, and behavioral processes using chromosome substitution strains. Intriguingly, epigenetic regulation of the *Y* chromosome may even be transgenerational, as a father's *Y* chromosome influences quantitative traits in his daughters, who do not carry *Y*'s, a result first noted in mice ((Nelson et al. 2010), and later confirmed in *Drosophila* (Friberg et al. 2012). The possibility of *Y* regulatory variation enhanced by *Y*-imprint in mammals is not unlikely, given that common epigenetic mechanisms regulate diverse imprinted domains in

animals, plants, and insects (reviewed in (MacDonald 2012)); however no known mammalian *Y* chromosome imprint has yet been observed.

As a cautionary note, the effects we see might not be ascribed only to parent-of-origin of the Y. One alternate explanation for the differential regulatory and PEV effects of the Y chromosome is that males with paternally versus maternally inherited Y's also had XX versus XXY mothers. We see that for PEV, this explanation is not satisfactory, as males showed differential expression of an X-linked w^{m4h} gene when the Y chromosome was inherited paternally or maternally (Supplementary Figure S2.5), even though ploidy of mothers remained constant (XXY in both cases). Position effect variegation and chromatin structure are closely related, since genes that influence PEV have been found to encode for structural components of chromatin, which in turn regulates the expression and repression of genes (Henikoff et al. 1992). Thus, although we cannot directly test for the effect of XXY aneuploidism on Y-regulatory effects, we believe maternal aneuploidism plays at most a small role in contrast to the large, demonstrated role of Y-imprinting via parent-of-origin transmission. A second explanation may be that the paternal transmission of the X exerts a strong influence, and that X-imprinting may lead to these results. This may be particularly relevant to the discussion of male versus female-based genes, as the X has been theorized to be a hot spot for sex-specific evolution (REF). If gene expression changes were the result of *cis*-regulatory control due to imprinting on the X, we would expect to see an over-representation of X-linked genes showing differential expression due to pattern of inheritance alone (normal versus reverse), irrespective of Y chromosomes. However, since we see no such pattern (genes localize to the X, 2^{nd}, 3^{rd} , and 4^{th} chromosomes at

12%, 49%, 38%, and < 1%, which does not differ from a random distribution based on gene densities, P = 0.139), we believe paternal transmission of the X cannot explain the pattern we see.

Our results support the hypothesis of *Y*-linked heterochromatic polymorphisms operating via chromatin remodeling to facilitate adaptive evolution of a suite of genes associated with male-specific functions. Such epigenetic regulation of gene expression, together with previous studies of the nontransitivity of sperm competition (Clark et al. 2000) and *Y*-by-background interactions for both male fitness (Chippindale and Rice 2001) and global gene expression (Jiang et al. 2010) highlight complex mechanisms for maintaining *Y*-polymorphisms responsible for regulating differential expressivity and variation in ecologically important traits.

WORKS CITED

Anaka M, Lynn A, McGinn P, Lloyd VK. 2009. Genomic imprinting in Drosophila has properties of both mammalian and insect imprinting. Dev. Genes Evol. 219:59–66.

Bachtrog D, Hom E, Wong KM, Maside X, de Jong P. 2008. Genomic degradation of a young Y chromosome in Drosophila miranda. Genome Biol 9:R30.

Baker BS, Belote JM. 1983. Sex determination and dosage compensation in Drosophila melanogaster. Annual Review of Genetics 17:345–393.

Bayes JJ, Malik HS. 2009. Altered heterochromatin binding by a hybrid sterility protein in Drosophila sibling species. Science 326:1538–1541.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological):289–300.

Brideau NJ, Flores HA, Wang J, Maheshwari S, Wang X, Barbash DA. 2006. Two Dobzhansky-Muller genes interact to cause hybrid lethality in Drosophila. Science 314:1292–1295.

Bull JJ. 1985. Evolution of Sex Determination Mechanisms. Menlo Park: Benjamin/Cummings

Carvalho AB, Koerich LB, Clark AG. 2009. Origin and evolution of Y chromosomes: Drosophila tales. Trends Genet 25:270–277.

Carvalho AB, Vaz SC, Klaczko LB. 1997. Polymorphism for Y-linked suppressors of sex-ratio in two natural populations of Drosophila mediopunctata. Genetics 146:891–902.

Carvalho AB. 2002. Origin and evolution of the Drosophila Y chromosome. Current Opinion in Genetics & Development 12:664–668.

Chakir M, Chafik A, Moreteau B, Gibert P, David J. 2002. Male sterility thermal thresholds in Drosophila: D-simulans appears more cold-adapted than its sibling D-melanogaster. Genetica 114:195–205.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philos Trans R Soc Lond, B, Biol Sci 355:1563–1572.

Chippindale A, Rice W. 2001. Y chromosome polymorphism is a strong determinant of male fitness in Drosophila melanogaster. Proc Natl Acad Sci USA 98:5677–5682.

Clark A, Dermitzakis E, Civetta A. 2000. Nontransitivity of sperm precedence in Drosophila. Evolution 54:1030–1035.

David J, Araripe L, Chakir M, Legout H, Lemos B, Petavy G, Rohmer C, Joly D, Moreteau B. 2005. Male sterility at extreme temperatures: a significant but neglected phenomenon for understanding Drosophila climatic adaptations. J Evol Biol 18:838–846.

Dimitri P, Pisano C. 1989. Position effect variegation in Drosophila melanogaster: relationship between suppression effect and the amount of Y chromosome. Genetics 122:793–800.

Dorer DR, Henikoff S. 1994. Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. Cell 77:993–1002.

Friberg U, Stewart AD, Rice WR. 2012. X- and Y-chromosome linked paternal effects on a life-history trait. Biol Lett 8:71–73.

Gibbons RJ, McDowell TL, Raman S, O'Rourke DM, Garrick D, Ayyub H, Higgs DR. 2000. Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nature Genetics 24:368–371.

Gnad F, Parsch J. 2006. Sebida: a database for the functional and evolutionary analysis of genes with sex-biased expression. Bioinformatics 22:2577–2579.

Golic KG, Golic MM, Pimpinelli S. 1998. Imprinted control of gene activity in Drosophila. Curr Biol 8:1273–1276.

Haller BS, Woodruff RC. 2000. Varied expression of a Y-linked P[w+] insert due to imprinting in Drosophila melanogaster. Genome 43:285–292.

Hearn MG, Hedrick A, Grigliatti TA, Wakimoto BT. 1991. The effect of modifiers of position-effect variegation on the variegation of heterochromatic genes of Drosophila melanogaster. Genetics 128:785–797.

Henikoff S, Dorer DR, Sabl JF, Jackson MH, Talbert PB. 1992. Position effect variegation and chromatin proteins. Bioessays [Internet] 14:605–612.

Huttunen S, Aspi J. 2003. Complex inheritance of male courtship song characters in Drosophila virilis. Behav Genet 33:17–24.

Innocenti P, Morrow EH. 2010. The sexually antagonistic genes of Drosophila melanogaster. Plos Biology 8:e1000335.

Jiang P-P, Hartl DL, Lemos B. 2010. Y not a dead end: epistatic interactions between Y-linked regulatory polymorphisms and genetic background affect global gene expression in Drosophila melanogaster. Genetics 186:109–118.

Joanis V, Lloyd VK. 2002. Genomic imprinting in Drosophila is maintained by the products of Suppressor of variegation and trithorax group, but not Polycomb group, genes. Mol Genet Genomics 268:103–112.

Krsticevic FJ, Santos HL, Januário S, Schrago CG, Carvalho AB. 2010. Functional copies of the Mst77F gene on the Y chromosome of Drosophila melanogaster. Genetics 184:295–307.

Larson K, Yan S-J, Tsurumi A, Liu J, Zhou J, Gaur K, Guo D, Eickbush TH, Li WX. 2012. Heterochromatin formation promotes longevity and represses ribosomal RNA synthesis. Plos Genet 8:e1002473.

Lemos B, Araripe L, Hartl D. Lemos, B et al Polymorphic Y supporting online material. Science.

Lemos B, Araripe LO, Hartl DL. 2008. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. Science 319:91–93.

Lemos B, Branco AT, Hartl DL. 2010. Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. PNAS 107:15826–15831.

Lindsley DL, Grell EH. 1969. Spermiogenesis without chromosomes in Drosophila melanogaster. Genetics 61:Suppl:69–Suppl:78.

Liu X, Wu B, Szary J, Kofoed EM, Schaufele F. 2007. Functional sequestration of transcription factor activity by repetitive DNA. J. Biol. Chem. 282:20868–20876.

Lloyd V. 2000. Parental imprinting in Drosophila. Genetica 109:35–44.

Lyckegaard EM, Clark A. 1989. Ribosomal DNA and Stellate gene copy number variation on the Y chromosome of *Drosophila melanogaster*. Proc Natl Acad Sci USA 86:1944–1948.

MacDonald WA. 2012. Epigenetic Mechanisms of Genomic Imprinting: Common Themes in the Regulation of Imprinted Regions in Mammals, Plants, and Insects. Genetics Research International 2012:1–17.

Maggert KA, Golic KG. 2002. The Y chromosome of Drosophila melanogaster exhibits chromosome-wide imprinting. Genetics 162:1245–1258.

Menon DU, Meller VH. 2009. Imprinting of the Y chromosome influences dosage compensation in roX1 roX2 Drosophila melanogaster. Genetics 183:811–820.

Migeon BR. 1998. Non-random X chromosome inactivation in mammalian cells. Cytogenet Cell Genet 80:142–148.

Montchamp-Moreau C, Ginhoux V, Atlan A. 2001. The Y chromosomes of Drosophila simulans are highly polymorphic for their ability to suppress sex-ratio drive. Evolution 55:728–737.

Moritz KB, Roth GE. 1976. Complexity of germline and somatic DNA in Ascaris. Nature 259:55–57.

Muller H. 1930. Types of visible variations induced by X-rays in Drosophila. Journal of Genetics 22:299–335.

Nelson VR, Spiezio SH, Nadeau JH. 2010. Transgenerational genetic effects of the paternal Y chromosome on daughters' phenotypes. Epigenomics 2:513–521.

Pal Bhadra M, Bhadra U, Birchler JA. 2006. Misregulation of sex-lethal and disruption of male-specific lethal complex localization in Drosophila species hybrids. Genetics 174:1151–1159.

Paredes S, Branco AT, Hartl DL, Maggert KA, Lemos B. 2011. Ribosomal DNA deletions modulate genome-wide gene expression: “rDNA-sensitive” genes and natural variation. Plos Genet 7:e1001376.

Paredes S, Maggert KA. 2009. Ribosomal DNA contributes to global chromatin regulation. Proc Natl Acad Sci USA 106:17829–17834.

Peng JC, Karpen GH. 2007. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. Nat. Cell Biol. 9:25–35.

Rice W. 1998. Male fitness increases when females are eliminated from gene pool: Implications for the Y chromosome. Proc Natl Acad Sci USA 95:6217–6221.

Rice WR, Gavrilets S, Friberg U. 2008. Sexually Antagonistic "Zygotic Drive‘’ of the Sex Chromosomes. Plos Genet 4:e1000313.

Rice WR. 1987. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics 116:161–167.

Rozen S, Marszalek JD, Alagappan RK, Skaletsky H, Page DC. 2009. Remarkably Little Variation in Proteins Encoded by the Y Chromosome's Single-Copy Genes, Implying Effective Purifying Selection. The American Journal of Human Genetics 85:923–928.

Sackton TB, Montenegro H, Hartl DL, Lemos B. 2011. Interspecific Y chromosome introgressions disrupt testis-specific gene expression and male reproductive phenotypes in Drosophila. PNAS 108:17046–17051.

Shen P, Wang F, Underhill P, et al. 2000. Population genetic implications from sequence variation in four Y chromosome genes. Proc Natl Acad Sci USA 97:7354–7359.

Singer JB, Hill AE, Burrage LC, et al. 2004. Genetic dissection of complex traits with chromosome substitution strains of mice. Science 304:445–448.

Smith C. 2007. The Release 5.1 Annotation of Drosophila melanogaster heterochromatin. Science 316:1586–1591.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology 3:Article3.

Weiler KS, Wakimoto BT. 1995. Heterochromatin and gene expression in Drosophila. Annual Review of Genetics 29:577–605.

Wu C-I, Ting C-T. 2004. Genes and speciation. Nature Reviews Genetics 5:114–122.

Xu B, Obata Y, Cao F, Taketo T. 2012. The Presence of the Y-Chromosome, Not the Absence of the Second X-Chromosome, Alters the mRNA Levels Stored in the Fully Grown XY Mouse Oocyte.Cooney AJ, editor. PLoS ONE 7:e40481.

Zhang P, Timakov B, Stankiewicz R, Turgut I. 2000. A trans-activator on the Drosophila Y chromosome regulates gene expression in the male germ line. Genetica 109:141–150.

Zhou J, Sackton TB, Martinsen L, Lemos B, Eickbush TH, Hartl DL. 2012. Y chromosome mediates ribosomal DNA silencing and modulates the chromatin state in Drosophila. PNAS.

Zuckerkandl E. 1974. A possible role of "inert" heterochromatin in cell differentiation. Action of and competition for “locking” molecules. Biochimie 56:937–954.

Zurovcova M, Eanes W. 1999. Lack of nucleotide polymorphism in the Y-linked sperm flagellar dynein gene Dhc-Yh3 of Drosophila melanogaster and D. simulans. Genetics 153:1709–1715.

Chapter 3

Accessible mutational trajectories for the evolution of pyrimethamine resistance in the malaria parasite *Plasmodium vivax*

Pan-Pan Jiang, Russell B. Corbett-Detig, Daniel L. Hartl, and Elena R. Lozovsky,

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge MA 02138

ABSTRACT

Antifolate antimalarials, such as pyrimethamine, have experienced a dramatic reduction in therapeutic efficacy as resistance has evolved in multiple malaria species. We present evidence from one such species, *Plasmodium vivax*, which has experienced sustained selection for pyrimethamine resistance at the dihydrofolate reductase (DHFR) locus since the 1970s. Using a transgenic *Saccharomyces cerevisiae* model expressing the *P. vivax* DHFR enzyme, we assayed growth rate and resistance of all 16 combinations of 4 DHFR amino acid substitutions. These substitutions were selected based on their known association with drug resistance, both in natural isolates and in laboratory settings, in the related malaria species *P. falciparum*. We observed a strong correlation between the resistance phenotypes for these 16 *P. vivax* alleles and previously observed resistance data for *P. falciparum*, which was surprising since nucleotide diversity levels and common polymorphic variants of DHFR differ between the two species. Similar results were observed when we expressed the *P. vivax* alleles in a transgenic bacterial system. This suggests common constraints on enzyme evolution in the orthologous DHFR proteins. The interplay of negative trade-offs between the evolution of novel resistance and compromised endogenous function varies at different drug dosages, and so too do the major trajectories for DHFR evolution. In simulations, it is only at very high drug dosages that the most resistant quadruple mutant DHFR allele is favored by selection. This is in agreement with common polymorphic DHFR data in *P. vivax*, from which this quadruple mutant is missing. We propose that clinical dosages of pyrimethamine may have historically been too low to select for the most resistant allele, or that the fitness cost

of the most resistant allele was untenable without a compensatory mutation elsewhere in the genome.

INTRODUCTION

The development of effective clinical therapeutics relies on an understanding of the molecular evolution of drug resistance. Factors which may affect evolutionary pathways of resistant alleles are drug concentration, epistatic interactions between mutations within the same gene, mutation bias in the genome, and adaptive conflict over resistance and function (Peterson, Walliker, and Wellems 1988; Snewin et al. 1989; Looareesuwan et al. 1996; Wellems 2002). The malaria parasite presents a unique opportunity for studying the constraints that a fitness landscape may impose on the evolution of an organism as it has historically undergone waves of strong selection for drug resistance. These waves have been marked by the introduction, and eventual compromised efficacy, of a changing arsenal of first-line malaria drugs, including chloroquine, atovaquone, and pyrimethamine (Breman 2012). Artemisenin-combination therapies (ACTs) are the last line of defense for the treatment of malaria, but signs of resistance building in small pockets of Southeast Asia demonstrates pervasive selection for resistance, and highlights the importance of understanding the molecular basis of resistance evolution (Kepler and Perelson 1998).

Drug concentration is considered critical for the selection of drug resistance in infectious agents such as HIV (White et al. 2009) and malaria (Kepler and Perelson 1998; Weinreich et al. 2006; Lozovsky et al. 2009; Brown et al. 2010; Chou et al. 2011). Varying concentrations of drugs have varying potential to facilitate resistance: there is no selection for resistance if drug concentrations are high enough to kill all sensitive and resistance parasites, or too low to kill either. Thus, the window of drug concentrations that allows evolution of resistant variants is relatively narrow (Peterson et al. 1988;

Snewin et al. 1989; Looareesuwan et al. 1996; Wellems 2002; Weinreich et al. 2006). A second factor to constrain evolutionary trajectories of resistant phenotypes is epistasis among multiple mutant sites in the same gene: the effect of any amino acid replacement may be dependent on the genetic context in which it finds itself, for example the same mutation may increase resistance on some backgrounds but decrease it in others. This type of non-additive interaction is termed sign epistasis, and has been reported in multiple instances, including beta-lactamase in *Escherichia coli* (Weinreich et al. 2006) and dihydrofolate reductase in *Plasmodium falciparum* (Brown et al. 2010). Another factor, mutation bias, can restrict the likelihood of fixation of a mutation (however favorable it may be) by limiting the likelihood of de-novo mutations to a particular base. For example, the genome of *P. falciparum* is strongly biased toward AT, and AT-rich codons are thus favored over others (Depristo, Zilversmit, and Hartl 2006). Finally, adaptive conflict can constrain evolutionary pathways because amino acid replacements which confer high resistance can come at a fitness cost by impairing endogenous functioning of the protein in the absence of drug (Sirawaraporn et al. 1997), although such negative trade-offs are not always necessary (Aharoni et al. 2005; Kondrashov 2005).

In this study, we focus on the malaria parasite *Plasmodium vivax* and the evolution within this species of drug resistance to pyrimethamine, a drug that competitively inhibits dihydrofolate reductase (DHFR). DHFR is a parasite enzyme required for the synthesis of tetrahydrofolate, an essential precursor of purines and several amino acids (Kompis, Islam, and Then 2005). Although *P. vivax* is not normally treated with pyrimethamine directly, the high frequency of mixed infections with *P. falciparum* (which has been treated by the combination drug sulfadoxine-pyrimethamine

since the 1970s), has exposed a large number of *P. vivax* parasites to selection for drug resistance in many areas. *P. vivax* is remotely related to *P. falciparum*, the two having diverged approximately 100 million years ago (Ayala, Escalante, and Rich 1999; Carter 2003). The life cycle of *P. vivax* features a latent liver infection responsible for relapses months or years following the initial infection. Although seldom fatal, *P. vivax* elicits incapacitating clinical symptoms and recurrent severe relapses. Recent work suggests that the morbidity and mortality associated with these cycles of reinfection are greater than previously believed (Mendis et al. 2001; Sina 2002).

The combination of a long history of pyrimethamine use and the common co-infection of *P. vivax* and *P. falciparum* has facilitated the evolution of drug resistance in both species. Point mutations in the *dhfr* gene and their impact on sensitivity of *P. falciparum* to pyrimethamine have been studied extensively in vitro and in vivo (Wooden et al. 1997; Cortese and Plowe 1998; Randrianasolo et al. 2004). However, similar data for *P. vivax* is lacking due to the difficulties associated with culturing this species in the laboratory.

The active site regions of DHFR in *P. vivax* and *P. falciparum* are strongly conserved, despite the two proteins preserving only ~66% sequence identity (Kongsaeree et al. 2005). The *dhfr* coding region is encoded by about 700 nucleotides in both species. However, work in *P. falciparum* shows that only a limited number of mutations are associated with drug resistance. Mutations at codons 16, 50, 51, 59, 108, and 164 have been observed to affect resistance in pyrimethamine, or the related antifolate drugs cycloguanil or chlorcycloguanil, in *P. falciparum* isolates from around the world (Sibley et al. 2001; Gregson and Plowe 2005; Hyde 2005). Four of these mutations (N51I, C59R,

S108N, and I164L) were found to be most commonly associated with drug resistance, and were extensively studied by Lozovsky et al. (Lozovsky et al. 2009) to predict favored pathways of drug resistance based on empirical estimates of mutational spectrum and probabilities of fixation based on relative levels of resistance. Although far fewer isolates of *P. vivax* have been studied, numerous alleles of *dhfr* have been identified, which is in contrast to the lack of diversity in *P. falciparum* (Hawkins et al. 2008).

Sequence alignment indicates that the four mutations important to pyrimethamine resistance in *P. falciparum* DHFR (N51I, C59R, S108N, and I164L) correspond to mutations N50I, S58R, S117N, and I173L in *P. vivax* DHFR. Interestingly, mutations at these exact codons have been observed in natural *P. vivax* isolates, suggesting that *P. falciparum* and *P. vivax* DHFR may experience similar functional and selective constraints despite approximately 100 million years of divergence. However, the story may not be so simple, as mutations which are common and convey high resistance in *P. falciparum*, such as I164L, are rarely observed in *P. vivax* (where the homologous mutation is I173L). Therefore, evolutionary trajectories of mutation acquisition by the two proteins may differ more than expected.

The goal of this experiment was to assay whether these four point mutations in DHFR, which in *P. falciparum* have been demonstrated to convey drug resistance and mold the evolutionary landscape of mutation acquisition, have similar effects in *P. vivax*. The mutational landscape in DHFR differs between *P. falciparum* and *P. vivax* quite extensively, with far greater polymorphism in the latter species. Some mutations which are commonly observed in *P. vivax,* which are not included in the set we have studied, include F57I, T61M, and S117T. Due to this, the emphasis of our study is not on practical

applications of the results in a clinical setting, but rather to test whether orthologous amino acid replacements in orthologous proteins have parallel effects. And to apply a forward-time approach (rather than traditional coalescent models) to simulate evolutionary trajectories that use fitness landscapes informed by multiple parameters such as drug concentration, direct selection coefficients, and drift (rather than IC50 alone).

Using a transgenic yeast system, we explored the mutational landscape of pyrimethamine resistance in *P. vivax* DHFR. Following others (Weinreich et al. 2006; Lozovsky et al. 2009; Brown et al. 2010) we genetically engineering all 16 possible combinations of 4 amino acid replacements. Each of these mutations has been observed in various combinations in nature and shown to be associated with drug resistance. Fitness landscapes and stepwise mutational trajectories can be predicted by analyzing the phenotypes of these alleles in the presence of the drug. Importantly, trade-offs between novel function (resistance) and existing function (activity) can be gauged simultaneously by translating growth rates at various drug concentrations into proportional fitness peaks and valleys in the resulting landscape.

Overall, our results provide the first predicted pathways of mutation acquisition in the *dhfr* gene in *P. vivax*. Importantly, these pathways are given for a gradient of drug concentrations, which can represent a temporally and spatially heterogeneous drug reservoir within the population created by sporadic noncompliance to antibiotic regimes, varying dosage regimes, or compartmentalization of drugs in different tissues. We show that varying drug concentrations affect the fitness of the most resistant alleles. In other words, adaptive fitness landscapes are strongly molded by drug concentrations. The landscape is dominated by the single peak of the non-mutant allele at the low end of drug

concentration spectrum, and is effectively flat (and without a single peak) for very high drug concentrations. Medium-range drug concentrations have distinctly different projected evolutionary trajectories, and the selectivity window for the most resistant quadruple mutant is constrained within a narrow range. Second, *P. vivax* DHFR shows sign epistasis at multiple mutation sites, which explains why many evolutionary pathways are rendered inaccessible to selection in our scenarios. Finally, to address the ability of adaptive conflict to influence evolution, we used a composite measure of resistance and endogenous function, both of which determine fitness, to represent the nodes in our fitness landscapes. This is in contrast to previous studies that primarily relied on resistance only to construct adaptive landscapes (Weinreich et al. 2006; Lozovsky et al. 2009; Brown et al. 2010).

METHODS AND MATERIALS

E.coli and yeast strain construction

P. vivax DHFR coding sequence was isolated and cloned into the vector pET17 (Novagen). The resulting plasmid was transformed into *E. coli* strain HMS174(DE3) (Novagen). All 16 possible combinations of mutations at 4 sites (N50I, S58R, S117N, I173L) were introduced by QuikChange Site-Directed Mutagenesis Kit (Strategene). Each mutageneized plasmid was sequenced to confirm the presence of expected mutations, and the absence of any other mutations. Plasmids containing all mutated alleles of DHFR were then introduced into *E.coli* strain LH18, thyA Δfol:kan donated by E.E. Howell (Howell, Foster, and Foster 1988). This strain lacks dihydrofolate reductase, as well as thymidylate synthase, and requires plating on full LB medium supplemented

with 200 mg/mL of thymine for growth. In addition, this strain carries a kanamycin-resistance selective marker. Once transformed, this strain can be grown on Bonner-Vogel minimum medium supplemented with 200 mg/mL of thymine.

To create yeast strains with the same mutated alleles, we cloned each allele into the GR7 shuttle vector, a derivative of the pRS314 yeast shuttle vector (Sikorski and Hieter 1989). We used *S. cerevisiae* strain TH5 (*MATa leu2-3,112 trp1 ura3-53, dhfr1::URA3 tup1*, provided by Carol Sibley) to assay pyrimethamine resistance conferred by each DHFR allele. TH5 lacks DFR1, the yeast orthologue of the DHFR gene, and when not transformed with functional DHFR, requires media supplemented with 100 ul/mL deoxythymidine monophosphate (dTMP, Sigma-Alrich) for growth. TH5 was transformed with each of the 16 alleles of DHFR on tryptophan drop-out medium (SC trp-) to select for the presence of the GR7 vector carrying DHFR. The SC trp- medium represents "minimal medium" for yeast strains in all future references.

We used both yeast and bacterial transgenic systems for our assays to rule out possible artifacts of background mutations in the vector to ensure that the results represented biologically relevant behavior of the parasite DHFR alleles themselves. In this paper, we focus on the results from the yeast transgenic system, but results from both systems are strongly correlated.

Growth Rate Calculations

For each strain, we picked two to five colonies from a solid media plate and inoculated the appropriate liquid minimum medium culture with our transformed *E.coli* and yeast strains. After overnight culture, cultures were diluted to an OD_{600} of 0.01 (or

approximately 6 x 10^4 cells/mL) in a series of concentrations of pyrimethamine in minimum medium, and then dispensed into microtiter plates. These plates were transferred to a Bioscreen C microbiological workstation (Thermo Labsystems), which recorded OD_{600} readings every 15 minutes for 2 days. Growth rate was calculated by taking a least squares linear regression of log absorbance versus time for a 3 hour sliding window over the length of the growth curve (Joseph and Hall 2004; Brown et al. 2010). Growth rates represent the maximum regression coefficient among all sliding windows.

IC50 Calculations

We calculated the resistance of each strain using inhibitory concentration 50 (IC50) measurements, which represent the pyrimethamine concentration at which growth rate is 50% of what it is in the absence of pyrimethamine. IC50 values were obtained as follows. For each strain, we fit the following logarithmic curve to our growth rate versus pyrimethamine concentration:

$$G_i = \frac{A_i}{1 + e^{\frac{x - b_i}{c_i}}}$$

where G_i is the growth rate of strain i, A_i is the maximum growth rate in the absence of pyrimethamine, x is the natural log of (pyrimethamine concentration + 1), b_i is the pyrimethamine concentration at which G_i is half of A_i, and c_i is a scaling parameter determining the shape of the logistic curve. This curve gave us predicted growth rates at a range of pyrimethamine concentrations. Nonlinear least squares regressions were used to determine the value of b_i, which represents the IC50 values for each strain.

Calculation of possible evolutionary trajectories

To investigate properties of the adaptive landscape, we conducted forward-time simulations using simuPOP (Peng and Kimmel 2005). Each simulation began with a population of individuals that were fixed for the ancestral haplotype. We drew the population size for each simulation from an $e^{unif(ln(1000),ln(100000)}$ distribution, or a natural-log scaled uniform distribution from 1,000 to 100,000 individuals. Mutations were then added according the relative rate matrix for *P. vivax* that we computed from the data presented in Neafsey *et al*. (Neafsey et al. 2012). To convert the relative substitution rate matrix to more realistic per-generation rate matrix, we divided all relative rates by a factor of 10^3. Ultimately, we are interested in the rate of amino acid substitutions; hence to compute the per-generation substitution rates, we summed the rates of all nucleotide substitutions that can produce the amino acid substitutions of interest. We performed 500 simulations at pyrimethamine concentrations 0, 2, 4, 6, 8 corresponding to the transformation ln(pyrimethamine concentration + 1). Concentrations above 8 were not explored because for most population sizes, substitutions are "nearly neutral", *i.e*. $Ns < 1$ (over the range of population sizes investigated), thus the adaptive landscape is effectively flat. Growth rate for each haplotype at each concentration was computed as explained above in “Growth Rate Calculations”. We then assigned selective coefficients to each haplotype based on the relative maximum growth rates estimated for each concentration. Simulations were run until the population fixed the haplotype of optimal fitness or until 1,000,000 generations had passed.

We used these simulations to determine the adaptive trajectory for each simulation as follows. We recorded the frequency of each haplotype every ten

generations during the simulation. If, at one time step, any haplotype had a frequency greater than 0.5, we considered the population to be in this state. We then recorded the transitions between haplotypes and removed loops (e.g. a transition from haplotype 3 to 4 and immediately back again to 3 is recorded as just 3) to determine the adaptive trajectory for that simulation. These were then compiled for each concentration and the four most likely adaptive trajectories determined from the aggregate.

RESULTS

In order to assess how various mutational combinations in the coding sequence of *P. vivax* DHFR affected resistance to pyrimethamine, a drug which competitively inhibits this protein's activity, we created all 16 possible combinations of the mutants N50I, S58R, S117N, and I173L. The mutagenized DHFR coding sequences were cloned into a yeast vector and transformed into a yeast strain lacking DFR1, the yeast orthologue of the DHFR gene.

We estimated the level of pyrimethamine resistance as the concentration of drug that inhibited cell growth by half, a metric known as the IC50 (see Methods). The results are shown in the barplot in Figure 3. 1. The DHFR alleles indicated along the horizontal axis are given in the form of a vector of 0's and 1's corresponding, from left to right, the amino acid residues 50, 58, 117, and 173. Each 0 indicates a nonmutant codon, and each 1 indicates a mutant codon. Therefore 0000 indicates the wildtype, nonmutated allele, and 1111 indicates the quadruple mutant. Figure 3. 1 shows that of the four single mutants, 0010 is much more resistant to pyrimethamine, and may likely be favored as the

first mutational step. Consistent with data from previous yeast or bacterial complementation systems for *P. falciparum* DHFR (Lozovsky et al. 2009; Brown et al. 2010), the quadruple mutant (N50I, S58R, S117N, I173L) was the most resistant, with triple mutants also exhibiting high levels of resistance. Clinical data from malarial patients treated with a sulfadoxine-pyrimethamine drug regimen showed higher failure rates with *P. vivax* DHFR quadruple mutants compared to wildtype (Tjitra et al. 2002). Figure 3. 2 shows estimated growth rates of all alleles across a range of drug concentrations. Alleles vary not only at the point at which they cross the Y-axis (growth in the absence of pyrimethamine), or concentration of drug at which growth is reduced by half (IC50), but also by the shape of the logistic growth curve (whether growth falls sharply with increasing dosage of drug). Beyond the obvious low resistance of the wildtype 0000 allele and the very high resistance of the quadruple mutant 1111 allele, the evolutionary landscape is complex. Epistatic interactions among the mutant sites are common (Table 3.1). For example, the S58R mutation increases resistance in 4 backgrounds, has a negligible effect on resistance in 3 backgrounds, and actually decreases resistance in 1 allelic background.

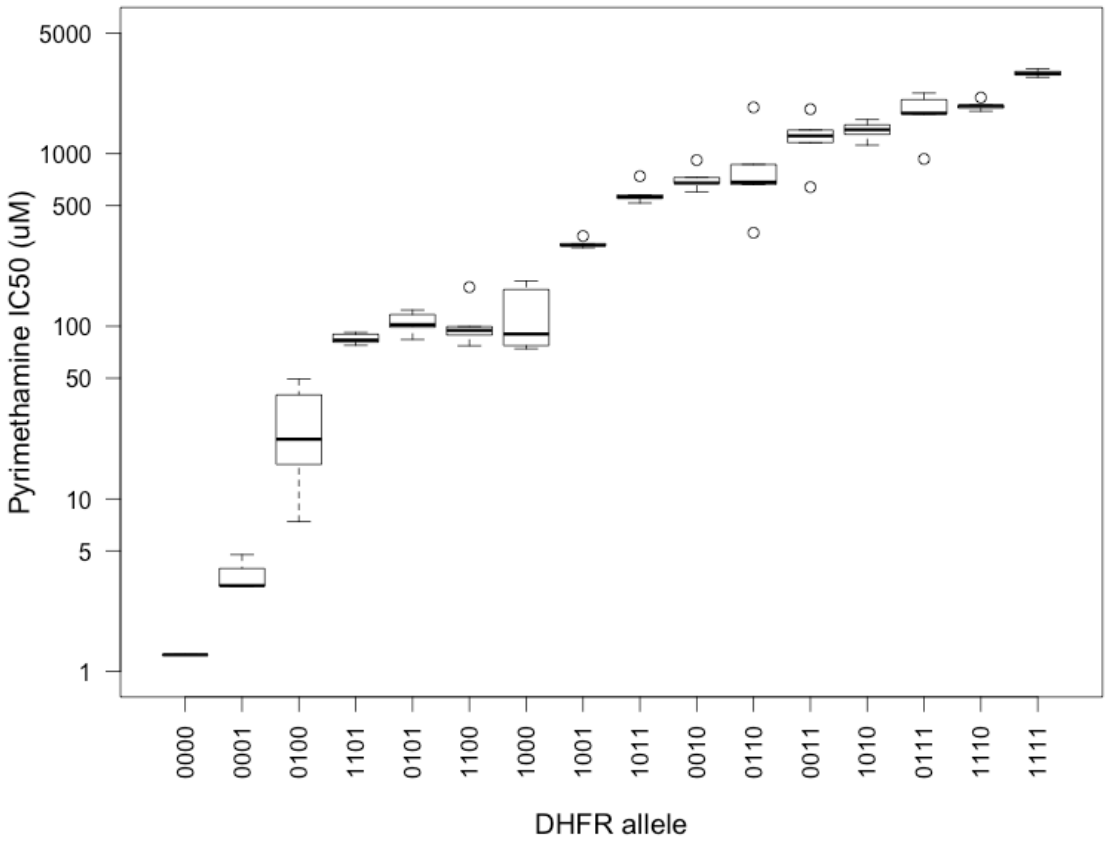


Figure 3.1. Boxplot showing resistance phenotypes of 16 DHFR alleles mutated at four possible sites. Alleles are shown in order of increasing resistance.

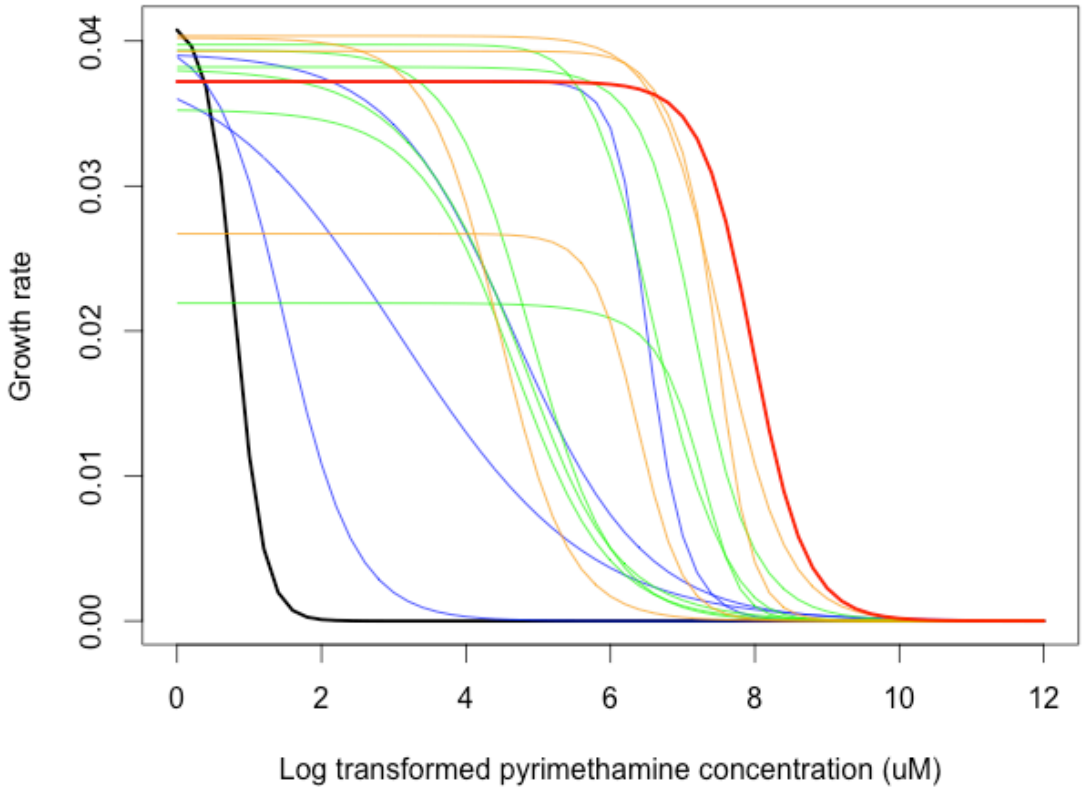


Figure 3.2. Fitted logarithmic growth rates of all alleles at increasing concentration of pyrimethamine. Concentration of pyrimethamine (x) is transformed as $\ln(x+1)$. Black line represents wildtype (0000) allele, which is least resistance to pyrimethamine. Single mutants (1000, 0100, 0010, 0001) are represented by blue lines. Double (0011, 0101, etc) and triple mutants (1110, 1011, etc) are represented by green and orange lines, respectively. Red line quadruple mutant (1111) is the most resistant allele.

Table 3.1. Summary of mutational effects on pyrimethamine resistance in DHFR. *Differences in mean IC50 values are significant at $P < 0.05$, after Bonferroni correction. Mean proportional increase of IC50, geometric mean across all 8 alleles.

Mutation	DHFR alleles on which mean mutational effect is			Mean fold proportional increase
	Positive*	Negative*	Negligible	
N50I	3	0	5	3.85
S58R	4	1	3	2.25
S117N	8	0	0	29.9
I173L	2	1	5	1.63

Correlation between IC50 values for *P. vivax* and *P. falciparum dhfr* alleles were strongly significant and positive (rho = 0.8607, *P* = 2.2e-16, Spearman rank correlation test, *P. falciparum* DHFR data from (Lozovsky et al. 2009), Supplementary Figure S3.1). Allele 0011 was removed from this calculation because this allele failed to grow in *P. falciparum* DHFR, most likely due to adaptive conflict between resistance to inhibition and maintenance of enzyme activity. In contrast, the *P. vivax* 0011 DHFR allele maintained functional enzyme activity, and had high resistance to pyrimethamine.

The development of a new function often comes at a cost of a previous function, thus proteins evolving for increased drug resistance are often thought to be under adaptive conflict for endogenous function. In our case, the evolution of resistance to pyrimethamine may compromise the catalytic activity of DHFR. However, this kind of trade-off may not be necessary, as was suggested by Brown et al, (Brown et al. 2010), who observed no association between resistance level and growth rate. In our data, we also see no clear association between these two phenotypes, as correlation analysis

reveals (Pearson's correlation between log-transformed IC50 and maximum growth rate in the absence of drug, $P = 0.3416$). However, maximum growth rate in the absence of drug and IC50 are only two metrics, which taken individually paint an incomplete picture of fitness. As Figure 3.2 shows, drug concentration has a strong effect on relative fitness of alleles, with landscapes dominated by the wildtype (0000) allele in the absence of drug (pyrimethamine concentration equal to 0) and very flat at high drug concentrations. The flat fitness landscape at high drug concentrations means there are no pronounced peaks or valleys, and the population is fixing alleles mainly by drift. Simulations for evolutionary pathways in such landscapes can give us little insight into the likely evolutionary pathway of new mutations, and were excluded from analysis. Gene-by-environment interactions are important, as moderate dosages of pyrimethamine have greater power to select for resistant alleles. This is because, in contrast to the flat landscapes at either dosage extreme, moderate dosages induce greater variation in fitness differences between the alleles. In addition, effective population size of the malaria species and mutation bias in the genome can also affect the path of evolution. In order to simulate an evolutionary pathway for DHFR evolution, we incorporated all these elements in a forward time simulation (see Methods).

The results of the simulation are displayed in Figure 3.3. In the absence of drug, the fitness landscape is dominated by the fitness peak of the wildtype (0000) allele.

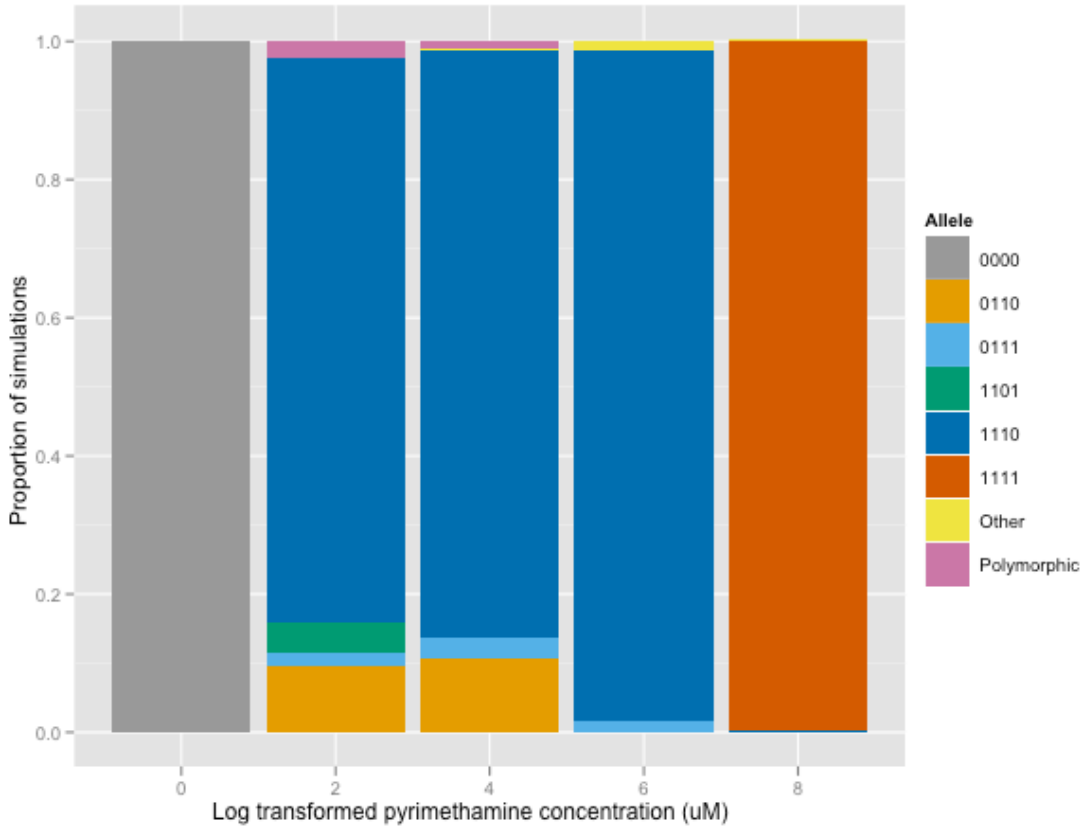


Figure 3.3. Proportion of simulation runs which fixed for any particular allele, at varying concentrations of pyrimethamine. Concentration of pyrimethamine (x) is transformed as $\ln(x+1)$. "Other" represents alleles which fixed, but with frequencies of less than 1%. "Polymorphic" represents simulations which remained segregating for polymorphic alleles at the end of 1 million generations.

Despite random mutations and sometimes small population sizes, no simulations ever wandered off this peak. At higher drug concentrations (log transformed pyrimethamine concentrations 2 to 6), the quadruple mutant does not have the highest fitness (Figure 3. 2), likely reflecting adaptive conflict between resistance and catalytic activity. The triple mutant 1110 allele fixes in the vast majority of cases, and populations that remain segregating for two or more polymorphic alleles at the end of 1 million generations make up a small percentage of the runs (2.4%, 1.0%, and 0% for log transformed concentrations 2, 4, and 6 respectively). At a log transformed pyrimethamine concentration of 8, which translates into an actual pyrimethamine concentration of ~3000 uM, the quadruple mutant 1111 allele fixes in 99.8% of simulated runs. At this concentration, the 1111 allele resides at the highest peak of the fitness landscape. Again, since we use growth rates to estimate fitness at different drug dosages, this fitness is informed by both resistance and catalytic activity. Taken together, these results suggest that despite being the most resistant allele, the quadruple mutant is not selectively favored unless very high dosages of drugs are administered. These simulations also reinforce the idea that effective population size plays a large role in the time to fixation of favorable alleles, with larger population sizes leading to faster fixation of favorable mutations (Figure S3.2). In addition, smaller effective population sizes were more likely to fix alleles which were not the absolute fitness peak in the landscape, because drift pushed them more frequently onto sub-optimal peaks (Table S3.1).

This result is consistent with amino acid replacements observed in worldwide surveys of *P. vivax* (Table 3.2). The majority of *P. vivax* samples isolated from patients in Southeast Asia and India were mutated in at least one of these four positions. In contrast

Table 3.2. Worldwide prevalence of polymorphic *P. vivax* DHFR alleles containing mutations N50I, S58R, S117N, and/or I173L. References in Table S3.3.

Allele	Location
S58R/S117N (0110)	Thailand, Myanmar, Cambodia, Myanmar, China, India, Mauritania, Iran, Pakistan, French Guiana, Ethiopia, Madagascar
S58R/S117N/I173L (0111)	French Guiana, Myanmar
S58R (0100)	Thailand, India, Indonesia, Vietnam, Vanuatu, China, East Timor
N50I/S58R/S117N/I173L (1111)	Not found

to the prevalence of the 1111 allele in *P. falciparum* samples (van den Broek et al. 2004; Anderson et al. 2005; Ahmed et al. 2006), this allele has not yet been observed in *P. vivax*, perhaps because clinical dosages of pyrimethamine never reach high enough levels for the quadruple mutant to have a fitness advantage. Another possibility is that long-term maintenance of the quadruple allele requires the presence of a compensatory mutation elsewhere in the genetic background, as has been shown in *P. falciparum* (Nair et al. 2008).

Other polymorphisms commonly found in natural isolates (such as the alleles 0110 and 0111) do coincide with several of the intermediates predicted by our simulated most likely pathways (Figure 3.4). Two things are notable from these results. First,

despite the striking similarity between IC50 of alleles between *P. falciparum* and *P. vivax*, the trajectories differ. For example, at the highest pyrimethamine concentration we examined, where the quadruple mutant fixed in most simulations, the first evolutionary step was most often 0100 in our simulations (Figure 3.4D), but 0010 in *P. falciparum* (Lozovsky et al. 2009). However, at lower pyrimethamine concentrations, 0010 was also the first evolutionary step in *P. vivax*. This highlights the subtle differences in fitness landscapes produced by using IC50 alone (as in the case of *P. falciparum* simulations) and by using a multi-parameter model incorporating drug concentration, selection, and drift, as we did with *P. vivax*. The second notable point is that the allele 1110, which is an important intermediate step in our evolutionary pathways, has not yet been isolated in natural populations. The overall increased nucleotide diversity in *P. vivax* compared to *P. falciparum* is reflected in the variation found in *P. vivax* DHFR alleles, where mutations other than the ones studied here are sometimes found at high frequencies. Epistatic interactions between these multitudes of mutations could result in evolutionary pathways that were unexplored in our simulations. For example, the 1110 background may stand as an unfavorable background for an otherwise highly favored mutation. Such sign epistasis is quite common, both within genes, as demonstrated by amino acid replacement S58R and I173L in our results (Table 3.2) and in β-lactamase in *E.coli* (Weinreich et al. 2006); as well as between genes, such as in *Methylobacterium* (Chou et al. 2011).

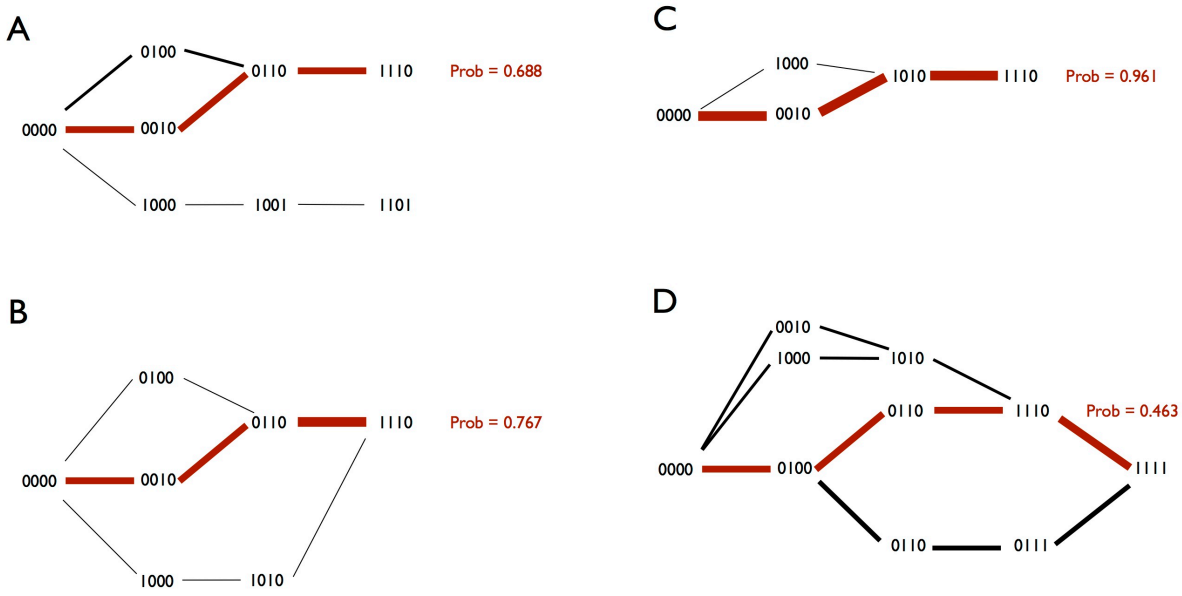


Figure 3.4. Preferred evolutionary pathways of pyrimethamine resistance in *P. vivax* DHFR. Panels A, B, C, and D represent log transformed pyrimethamine concentrations of $x = 2, 4, 6, 8$, respectively, where the transformation is evaluated at $\ln(x+1)$. The top four major pathways for each concentration are shown, except when probabilities fall below 0.01. Widths of lines in pathways correspond to their probabilities. The major pathway is shown in red, and given with an estimated probability.

DISCUSSION

We describe the resistance of mutant DHFR proteins in *P. vivax* to the anti-malarial drug pyrimethamine. Correlation analysis reveals that the adaptive landscape of the *P. vivax* DHFR alleles is highly correlated with the adaptive landscape of the *P. falciparum* DHFR alleles (Lozovsky et al. 2009). This suggests that orthologous mutations in the active sites of related proteins have similar functional significance for relatively distantly related species. This hypothesis is supported by more recent computational analysis of binding between malaria DHFR and anti-folates in four *Plasmodium* species – *P. falciparum, P. vivax, P. malariae,* and *P. ovale* – which found binding to be broadly similar, and determined by an analogous set of seven residues (Choowongkomon et al. 2010). In a way, this result is surprising, since *P. vivax* nucleotide diversity far surpasses that of *P. falciparum* nucleotide diversity, and mutations which are rare or missing in *P. falciparum* have been found to be important for resistance in *P. vivax* (Hawkins et al. 2007). Thus, we would not necessarily expect analogous mutations to have similar functional consequences. However, this result is also intuitively congruent with our current understanding of function in orthologous proteins that share common active sites.

The evolutionary pathways we identified to be important in the evolution of DHFR in *P. vivax* have been previously identified by in vivo studies of therapeutic efficacies of ACT-pyrimethamine combination drugs: natural isolates from Indonesia found mutations at residues 58 and 117 common in *P. vivax* isolates of malaria patients, as well as quadruple mutations (at residues 57, 58, 61, 117) (Tjitra et al. 2002). These

mutations were associated with increased resistance to sulfadoxine-pyrimethamine treatments. Based on these results, the authors suggested stepwise drug selection process in mutations favored mutation at residue 117 first, followed by mutations at 50, 58, which aligns with our results. Additional mutations favored by natural selection, but which were not assessed by us, included mutations at residues 57 and 61.

The serine to asparagine mutation in codon 117 (which corresponds to position 108 in *P. falciparum*) has repeatedly been demonstrated to be a major determinant of antifolate resistance (de Pécoulas et al. 1998; Sibley et al. 2001; Tjitra et al. 2002). The importance of the S117N mutation for resistance in *P. vivax* is even more stark than the S108N mutation in *P. falciparum*: in *P. vivax*, the S117N mutation confers a ~4000 fold increased resistance to pyrimethamine (Leartsakulpanich et al. 2002), whereas in *P. falciparum,* the S108N mutation confers only a ~100 fold increased resistance (Cowman et al. 1988; Peterson, Walliker, and Wellems 1988). Analysis of crystal structures of the DHFR protein reveals that a steric conflict arising from the side chain of a S117N mutant enzyme, accompanied by loss of binding to the serine at residue 120 is mainly responsible for the reduction in binding of pyrimethamine (Kongsaeree et al. 2005).

In addition to amino acid replacement S117N, in vitro assays of pyrimethamine sensitivity in mutated *P. vivax* isolates also identified S58R to be important for resistance (de Pécoulas et al. 1998). The importance of these mutations for resistance in *Plasmodium* DHFR has also been observed in *P. malariae,* where natural isolates resistant to pyrimethamine contained mutations S58R and S114N, corresponding to S58R and I173L in *P. vivax* and C59R and S108N in *P. falciparum* (Khim et al. 2012).

Some additional mutations which are associated with pyrimethamine resistance in *P. vivax* include F57L and T61M. Two amino acid replacements which have been found to be important for evolution of resistance in *P. falciparum* DHFR (N51I and I164L), are analogous to mutations N50I and I173L in *P. vivax,* but were very rarely observed in natural isolates in the latter species.

Pyrimethamine binds directly to the active site of DHFR, and competitively prevents the binding of dihydrofolates (Yuvaniyama et al. 2003). DHFR evolves resistance to pyrimethamine through acquiring substrate specificity for dihydrofolate (Rastelli et al. 2000). Such modifications to the active site of an enzyme is thought to impose greater trade-offs in native enzyme function than drugs that bind externally (Berkhout 1999; Tawfik 2005). Kinetic parameters of protein function are the pyrimethamine dissociation constant (K_i) and the catalytic turnover rate (k_{cat}). The relative fitness of parasites is highly informed by these two parameters, the first of which indicates level of resistance while the second indicates maintenance of endogenous function. In the case of malaria DHFR, adaptive conflict is in play as improved substrate specificity (K_i) often comes at the expense of catalytic activity (k_{cat}) (Sirawaraporn et al. 1997; Lozovsky et al. 2009) In other cases, the trade offs may be absent, as (Brown et al. 2010) showed that DHFR proteins are capable of evolving resistance without compromising existing function. We directly analyzed how adaptive conflict can shape evolution at various drug concentrations by using relative growth rates (which incorporate both substrate specificity and catalytic capacity) into our simulations of evolutionary pathways across fitness landscapes.

Our simulations showed that fitness landscapes could vary dramatically across drug concentrations. The wildtype allele (0000) shows the highest fitness in a drug-free environment, supporting the adaptive conflict theory that trade-offs can occur between resistance and growth. Although the quadruple mutant (1111) has the highest IC50 value, it is not the most favored allele at most drug concentrations. In fact, at lower drug concentrations, endogenous enzyme activity is compromised to an extent that limits the fitness benefits derived from higher drug resistance. A small window of drug concentrations can favor the fixation of the quadruple mutant, which in our assays was around 3000 uM pyrimethamine. However, it is difficult to predict how this concentration translates into an analogous dosage in a human patient, outside of a yeast transgenic system. Nevertheless, the incorporation of drug concentrations as a factor for predicted evolutionary pathways is a novel approach in tracking drug resistance in malaria research. This has important ramifications for human health. Since fitness landscapes are so strongly shaped by drug dosages, clinicians could modify dosage procedures over the course of treatment to subvert the evolution of resistance. This could be done perhaps by drawing populations into repeated cycles of suboptimal fitness using varying concentrations of drug, which induce previously highly fit alleles to reside at suboptimal fitness peaks in the current dosage regime.

Mathematical modeling based on another anti-malarial drug, mefloquine, suggest that drug dosage plays a strong role in selection of resistance (Simpson et al. 2000). They suggest that the initial deployment of lower doses provides an opportunity for selection of resistant alleles. This resistance would spread more rapidly than the de novo application of maximal doses. Our simulations show that maximal doses that do not allow the growth

of any alleles can circumvent resistance (in our yeast vector system, this does was ~3000 uM), however such high doses are probably eschewed as clinical treatments because of intolerable toxic side effects. Lower dosages of drugs can achieve adequate levels of parasetemia clearance with unmutated alleles, but also provides a breeding ground for resistance.

Ranked IC50 values for *S. cerevisiae* and *E. coli* strains carrying the same DHFR alleles are significantly correlated (rho = 0.7323, *P*-value = 0.001812, Table S3.2). The similarity of resistance phenotypes in two different species rules out experimental artifacts such as mutations in the genetic background of the yeast strains. Despite these similarities, the yeast and bacterial systems differ in some notable respects. For example, in yeast, a higher concentration of pyrimethamine is needed to inhibit DHFR activity sufficiently to measure IC50 than in bacterial cells: mean ln(IC50+1) values for *E.coli* is 3.58, versus 5.40 for *S. cerevisiae*. This may reflect a lower requirement for DHFR activity in yeast relative *to E. coli*, as has been previously noted (Brown et al. 2010).

An in-depth understanding of the molecular pathways of resistance evolution is necessary because infections by *P. vivax* pose a serious challenge to global health in Asia, South America, Central America, the Middle East, and parts of Africa (Hawkins et al. 2007; Parekh and Moorthy 2011). Our results lend novel insights into the evolution of anti-malarial resistant in the DHFR protein of *P. vivax*. The larger than expected agreement between major pathways identified in our analysis, and those observed in natural populations of *P. vivax* (Table 3.1), supports the use of model organisms in studies of drug resistance, particularly in organisms, such as *P. vivax,* that do not lend themselves to easy continuous culture in the laboratory. However, differences between

the laboratory setting and the field are manifold, and may explain why the most resistant quadruple mutant has not yet been found in nature. For example, we can create artificially high drug dosages in the laboratory that are avoided in the clinic, leading us to over-estimate the fitness of very resistant alleles, which likely evolve in a naturally lower drug dosage environment where their fitness advantage is reduced by negative trade-offs over endogenous function. Nonetheless, two results strengthen the validity of using model systems to predict evolutionary examples. The first is the congruence of growth rates for *P. vivax* alleles expressed in the *E. coli* and *S. cerevisiae* systems, which seems to rule out experimental artifacts. And the second is the extreme correlation between the fitness of corresponding alleles in *P. vivax* and *P. falciparum*. Both these observations strongly support the use of model organisms as a helpful system for studying the evolution of the malaria parasite.

WORKS CITED LIST

Aharoni A, Gaidukov L, Khersonsky O, McQ Gould S, Roodveldt C, Tawfik DS. 2005. The "evolvability" of promiscuous protein functions. Nature Genetics 37:73–76.

Ahmed A, Das MK, Dev V, Saifi MA, Wajihullah, Sharma YD. 2006. Quadruple mutations in dihydrofolate reductase of Plasmodium falciparum isolates from Car Nicobar Island, India. Antimicrob Agents Ch 50:1546–1549.

Anderson TJC, Nair S, Sudimack D, et al. 2005. Geographical distribution of selected and putatively neutral SNPs in Southeast Asian malaria parasites. Mol Biol Evol 22:2362–2374.

Ayala FJ, Escalante AA, Rich SM. 1999. Evolution of Plasmodium and the recent origin of the world populations of Plasmodium falciparum. Parassitologia 41:55–68.

Berkhout B. 1999. HIV-1 evolution under pressure of protease inhibitors: climbing the stairs of viral fitness. J. Biomed. Sci. 6:298–305.

Breman JG. 2012. Resistance to artemisinin-based combination therapy. Lancet Infect Dis 12:820–822.

Brown KM, Costanzo MS, Xu W, Roy S, Lozovsky ER, Hartl DL. 2010. Compensatory mutations restore fitness during the evolution of dihydrofolate reductase. Mol Biol Evol 27:2682–2690.

Carter R. 2003. Speculations on the origins of Plasmodium vivax malaria. Trends Parasitol 19:214–219.

Choowongkomon K, Theppabutr S, Songtawee N, Day NPJ, White NJ, Woodrow CJ, Imwong M. 2010. Computational analysis of binding between malarial dihydrofolate reductases and anti-folates. Malaria J 9:65.

Chou H-H, Chiu H-C, Delaney NF, Segrè D, Marx CJ. 2011. Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science 332:1190–1192.

Cortese JF, Plowe CV. 1998. Antifolate resistance due to new and known Plasmodium falciparum dihydrofolate reductase mutations expressed in yeast. Mol Biochem Parasitol 94:205–214.

Cowman AF, Morry MJ, Biggs BA, Cross GA, Foote SJ. 1988. Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of Plasmodium falciparum. Proc Natl Acad Sci USA 85:9109–9113.

de Pécoulas PE, Tahar R, Ouatas T, Mazabraud A, Basco LK. 1998. Sequence variations in the Plasmodium vivax dihydrofolate reductase-thymidylate synthase gene and their relationship with pyrimethamine resistance. Mol Biochem Parasitol 92:265–273.

Depristo MA, Zilversmit MM, Hartl DL. 2006. On the abundance, amino acid composition, and evolutionary dynamics of low-complexity regions in proteins. Gene 378:19–30.

Gregson A, Plowe CV. 2005. Mechanisms of resistance of malaria parasites to antifolates. Pharmacol. Rev. 57:117–145.

Hawkins VN, Auliff A, Prajapati SK, et al. 2008. Multiple origins of resistance-conferring mutations in Plasmodium vivax dihydrofolate reductase. Malaria J 7:72.

Hawkins VN, Joshi H, Rungsihirunrat K, Na-Bangchang K, Sibley CH. 2007. Antifolates can have a role in the treatment of Plasmodium vivax. Trends Parasitol 23:213–222.

Howell EE, Foster PG, Foster LM. 1988. Construction of a dihydrofolate reductase-deficient mutant of Escherichia coli by gene replacement. J Bacteriol 170:3040–3045.

Hyde JE. 2005. Drug-resistant malaria. Trends Parasitol 21:494–498.

Joseph SB, Hall DW. 2004. Spontaneous mutations in diploid Saccharomyces cerevisiae: more beneficial than expected. Genetics 168:1817–1825.

Kepler TB, Perelson AS. 1998. Drug concentration heterogeneity facilitates the evolution of drug resistance. Proc Natl Acad Sci USA 95:11514–11519.

Khim N, Kim S, Bouchier C, et al. 2012. Reduced Impact of Pyrimethamine Drug Pressure on Plasmodium malariae Dihydrofolate Reductase Gene. Antimicrobial agents ….

Kompis IM, Islam K, Then RL. 2005. DNA and RNA synthesis: antifolates. Chem. Rev. 105:593–620.

Kondrashov FA. 2005. In search of the limits of evolution. Nature Genetics 37:9–10.

Kongsaeree P, Khongsuk P, Leartsakulpanich U, Chitnumsub P, Tarnchompoo B, Walkinshaw MD, Yuthavong Y. 2005. Crystal structure of dihydrofolate reductase from Plasmodium vivax: pyrimethamine displacement linked with mutation-induced resistance. Proc Natl Acad Sci USA 102:13046–13051.

Leartsakulpanich U, Imwong M, Pukrittayakamee S, White NJ, Snounou G, Sirawaraporn W, Yuthavong Y. 2002. Molecular characterization of dihydrofolate reductase in relation to antifolate resistance in Plasmodium vivax. Mol Biochem Parasitol 119:63–73.

Looareesuwan S, Viravan C, Webster HK, Kyle DE, Hutchinson DB, Canfield CJ. 1996. Clinical studies of atovaquone, alone or in combination with other antimalarial drugs, for treatment of acute uncomplicated malaria in Thailand. Am J Trop Med Hyg 54:62–66.

Lozovsky ER, Chookajorn T, Brown KM, Imwong M, Shaw PJ, Kamchonwongpaisan S, Neafsey DE, Weinreich DM, Hartl DL. 2009. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. Proc Natl Acad Sci USA 106:12025–12030.

Mendis K, Sina BJ, Marchesini P, Carter R. 2001. The neglected burden of Plasmodium vivax malaria. Am J Trop Med Hyg 64:97–106.

Nair S, Miller B, Barends M, et al. 2008. Adaptive copy number evolution in malaria parasites. Plos Genet 4:e1000243.

Neafsey DE, Galinsky K, Jiang RHY, et al. 2012. The malaria parasite Plasmodium vivax exhibits greater genetic diversity than Plasmodium falciparum. Nature Genetics 44:1046–1050.

Parekh F, Moorthy V. 2011. Plasmodium vivax Vaccine Research: Insights from Colombian Studies. American Journal of Tropical Medicine and Hygiene 84:1.

Peng B, Kimmel M. 2005. simuPOP: a forward-time population genetics simulation

environment. Bioinformatics 21:3686–3687.

Peterson DS, Walliker D, Wellems TE. 1988. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. Proc Natl Acad Sci USA 85:9114–9118.

Randrianasolo L, Randriamanantena A, Ranarivelo L, Ratsimbasoa A, Domarle O, Randrianarivelojosia M. 2004. Monitoring susceptibility to sulfadoxine-pyrimethamine among cases of uncomplicated, Plasmodium falciparum malaria in Saharevo, Madagascar. Ann Trop Med Parasit 98:551–554.

Rastelli G, Sirawaraporn W, Sompornpisut P, Vilaivan T, Kamchonwongpaisan S, Quarrell R, Lowe G, Thebtaranonth Y, Yuthavong Y. 2000. Interaction of pyrimethamine, cycloguanil, WR99210 and their analogues with Plasmodium falciparum dihydrofolate reductase: structural basis of antifolate resistance. Bioorg. Med. Chem. 8:1117–1128.

Sibley CH, Hyde JE, Sims PF, et al. 2001. Pyrimethamine-sulfadoxine resistance in Plasmodium falciparum: what next? Trends Parasitol 17:582–588.

Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics 122:19–27.

Simpson JA, Watkins E, Price RN, Aarons L, Kyle DE, White NJ. 2000. Mefloquine Pharmacokinetic-Pharmacodynamic Models: Implications for Dosing and Resistance. Antimicrob Agents Ch 44:3414.

Sina B. 2002. Focus on Plasmodium vivax. Trends Parasitol 18:287–289.

Sirawaraporn W, Sathitkul T, Sirawaraporn R, Yuthavong Y, Santi DV. 1997. Antifolate-resistant mutants of Plasmodium falciparum dihydrofolate reductase. Proc Natl Acad Sci USA 94:1124–1129.

Snewin VA, England SM, Sims PF, Hyde JE. 1989. Characterisation of the dihydrofolate reductase-thymidylate synthetase gene from human malaria parasites highly resistant to pyrimethamine. Gene 76:41–52.

Tawfik DS. 2005. The evolvability of drug resistance: response to Fernandez. J Biomol Struct Dyn.

Tjitra E, Baker J, Suprianto S, Cheng Q, Anstey NM. 2002. Therapeutic Efficacies of Artesunate-Sulfadoxine-Pyrimethamine and Chloroquine-Sulfadoxine-Pyrimethamine in Vivax Malaria Pilot Studies: Relationship to Plasmodium vivax dhfr Mutations. Antimicrob Agents Ch 46:3947–3953.

van den Broek IVF, van der Wardt S, Talukder L, Chakma S, Brockman A, Nair S, Anderson TC. 2004. Drug resistance in Plasmodium falciparum from the Chittagong Hill Tracts, Bangladesh. Trop Med Int Health 9:680–687.

Weinreich DM, Delaney NF, Depristo MA, Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312:111–114.

Wellems TE. 2002. Plasmodium chloroquine resistance and the search for a replacement antimalarial drug. Science 298:124–126.

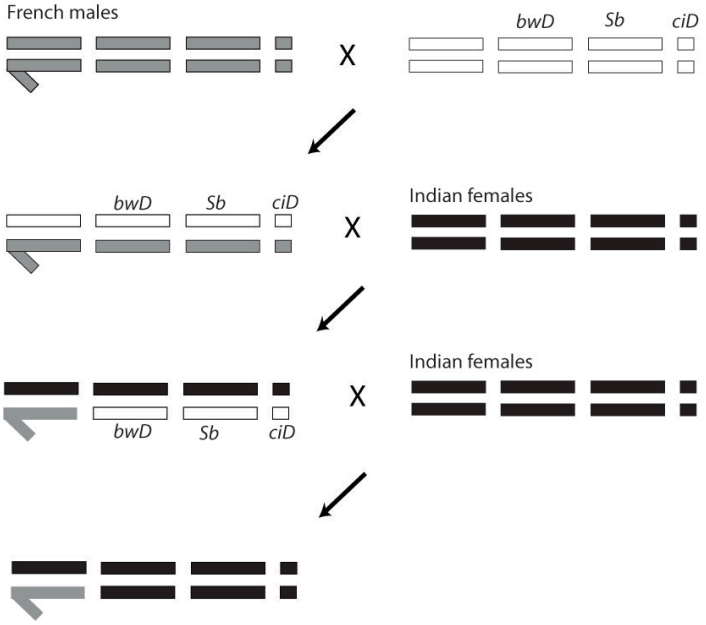
White NJ, Pongtavornpinyo W, Maude RJ, et al. 2009. Hyperparasitaemia and low dosing are an important source of anti-malarial drug resistance. Malaria J 8:253.

Wooden JM, Hartwell LH, Vasquez B, Sibley CH. 1997. Analysis in yeast of antimalaria drugs that target the dihydrofolate reductase of Plasmodium falciparum. Mol Biochem Parasitol 85:25–40.

Yuvaniyama J, Chitnumsub P, Kamchonwongpaisan S, Vanichtanankul J, Sirawaraporn W, Taylor P, Walkinshaw M, Yuthavong Y. 2003. Insights into antifolate resistance from malarial DHFR-TS structures. Nat Struct Biol 10:357–365.

SUPPLEMENTARY MATERIAL

A.



B.

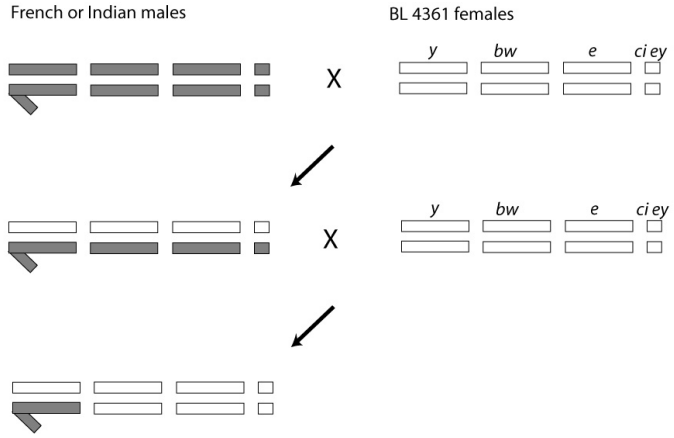


Figure S1.1 – A) Crossing scheme for substituting *Y* chromosomes from one population into the genetic background of the other. Drosophila males lack meiotic recombination, therefore no balancer chromosomes were necessary. Only crosses resulting in Y_F in an Indian genetic background are shown, but crosses resulting in Y_I in a French genetic background are analogous. B) Crossing scheme for introgressing *Y* chromosomes into a common laboratory stock genetic background.

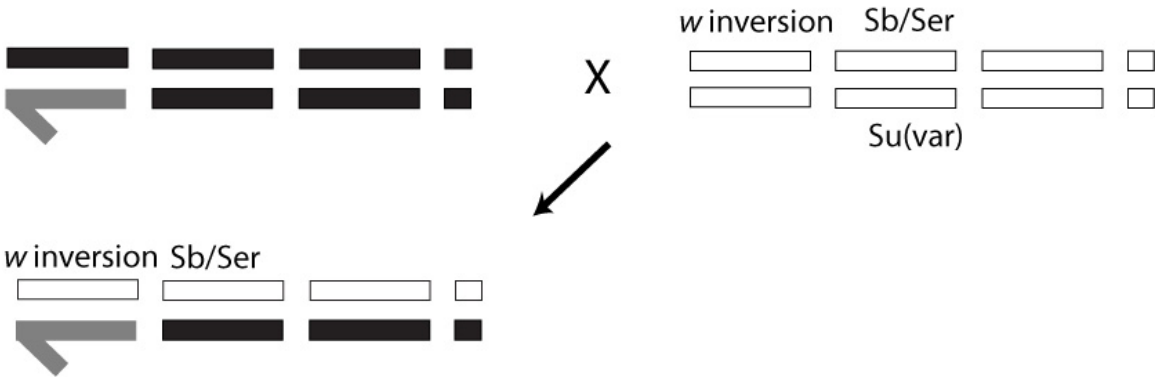


Figure S1.2 – Crosses to observe effects of *Y* chromosome and genetic background on position effect variegation. Bloomington Stock B6175 *(In(1)w[m4h]; Su(var)3-10[2]/TM3, Sb[1] Ser[1]*) females were used. Females have *X* chromosomes with an inversion placing *w[m4h]* close to the centromere. The second chromosome is heterozygous for a *Sb/Ser* dominant marker, selection of which removes the dominant *Su(var) (Suppressor of variegation)* and allows variegation of *w[m4h]*.

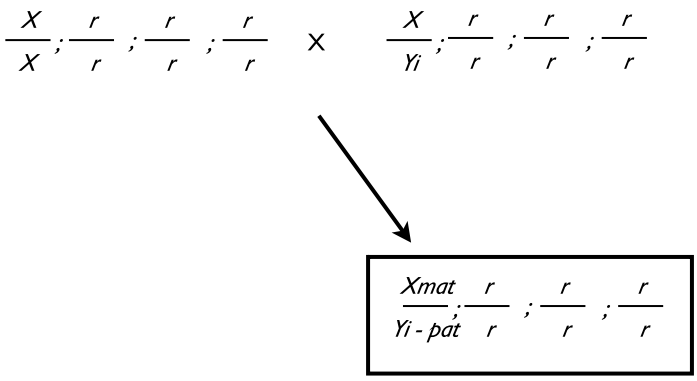
Table S1.1 – Number of genes identified by Maanova to be significantly modulated by *Y*-by-background effects (FDR < 0.01) which were also identified by BAGEL ($P > 0.95$) as differentially modulated by the *Y* chromosome in one, two, or all three of the genetic backgrounds examined (Indian, French, or B4361). The number of genes uniquely identified as displaying *Y*-by-background effects but not in the analysis with BAGEL are also shown.

	Number of genes overlapping with BAGEL (%)			
Maanova	No overlap	1 background	2 backgrounds	3 backgrounds
Y-by-background effects (346 genes in total	94 (27%)	169 (49%)	76 (22%)	7 (2%)

SUPPLEMENTARY MATERIALS

Figure S2.1. Crossing scheme for substituting creating males with A) normal sex chromosome inheritance (X-maternal, Y-paternal), or B) males with reverse sex chromosome inheritance (X-paternal, Y-maternal). Line 1: Attached X females were collected from Bloomington Stock 995, C(1;Y)3, ln(1)FM7, w[1] m[2]/0/C(1)M4, y[2]. Males carrying balancers on the 2nd and 3rd chromosomes were collected from Bloomington Stock 7199, w[*]; Kr[IF-1]/CyO; D[1]/TM6C, Sb[1] Tb[1]. Line 2 and 3: Yi males were collected from Y-substitution lines carrying Y-Cs or Y-Con. Crosses for these Y-substitution lines were carried out as previously described (LEMOS *et al.* 2008). Crosses were done independently for each Y chromosome lineage such that males have different Y chromosomes introgressed into a common isogenic background, with recessive markers (r) on each chromosome (these markers are y; bw; e; ci ey on the X, 1st, 2nd, 3rd, and 4th chromosomes, respectively).

A.



B.

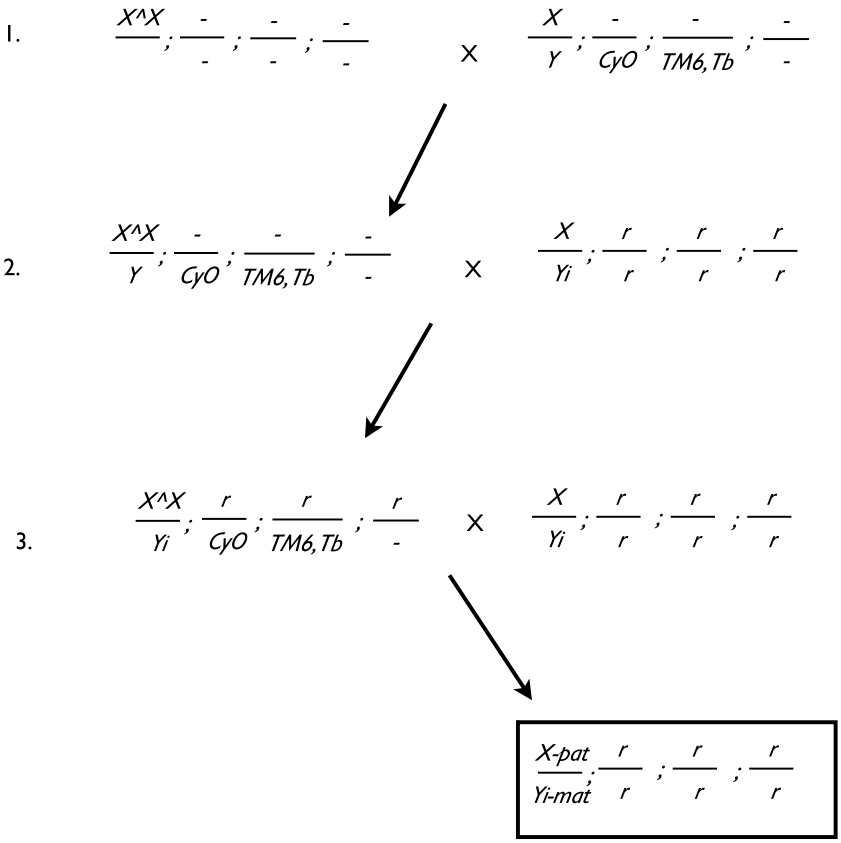


Figure S2.1. (Continued).

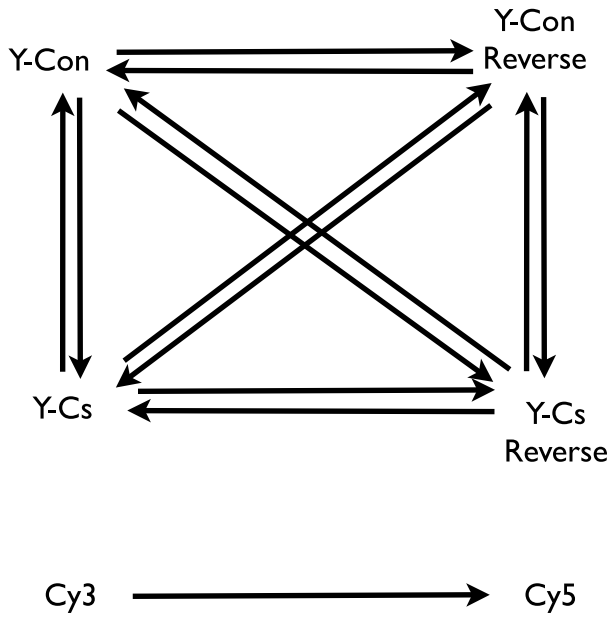
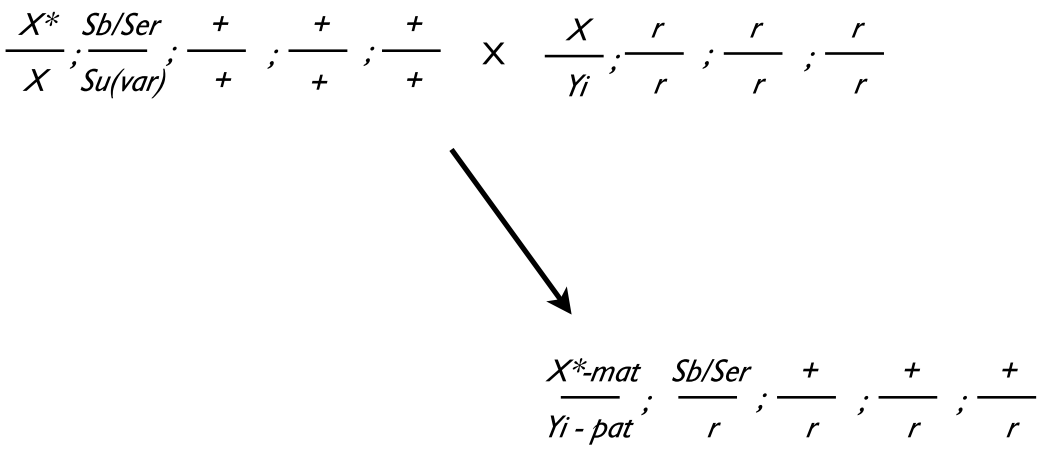


Figure S2.2. Experimental design of hybridization scheme for microarrays to collect genome-wide gene expression data. Each arrow denotes one hybridization reaction, with Cy3 labeling at the foot, and Cy5 labeling at the head, of the arrow.

Figure S2.3. Crosses to observe effects of reverse sex chromosome inheritance on position effect variegation (PEV). Bloomington Stock B6175 *(In(1)w[m4h]; Su(var)3-10[2]/TM3, Sb[1] Ser[1]*) females were used. Males and females have *X** chromosomes with an inversion placing *w[m4h]* close to the centromere. The second chromosome is heterozygous for a *Sb/Ser* dominant marker, selection of which removes the dominant *Su(var) (Suppressor of variegation)* and allows variegation of *w[m4h]*. A) Deriving males with normal sex chromosome inheritance, and variegating expression of *w[m4h]*. B) Deriving males with reversed sex chromosome inheritance, and variegating expression of *w[m4h]*.

A.



B.

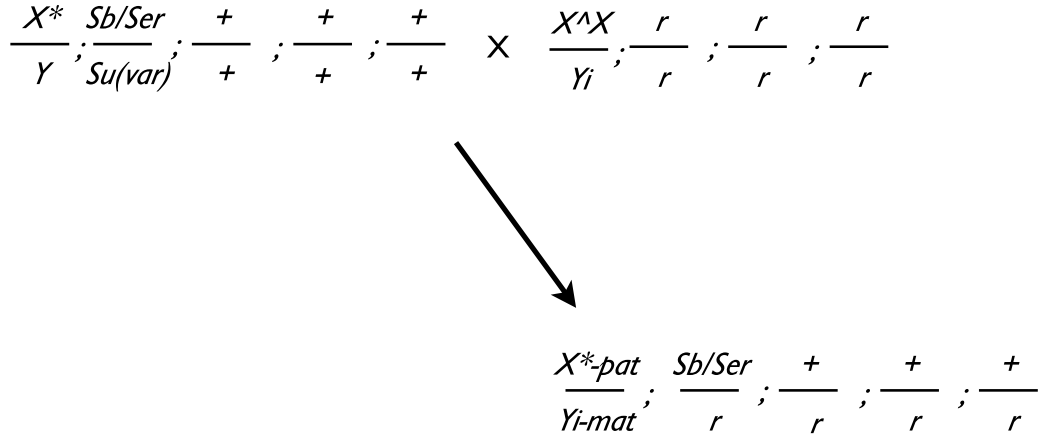


Figure S2.3. (Continued).

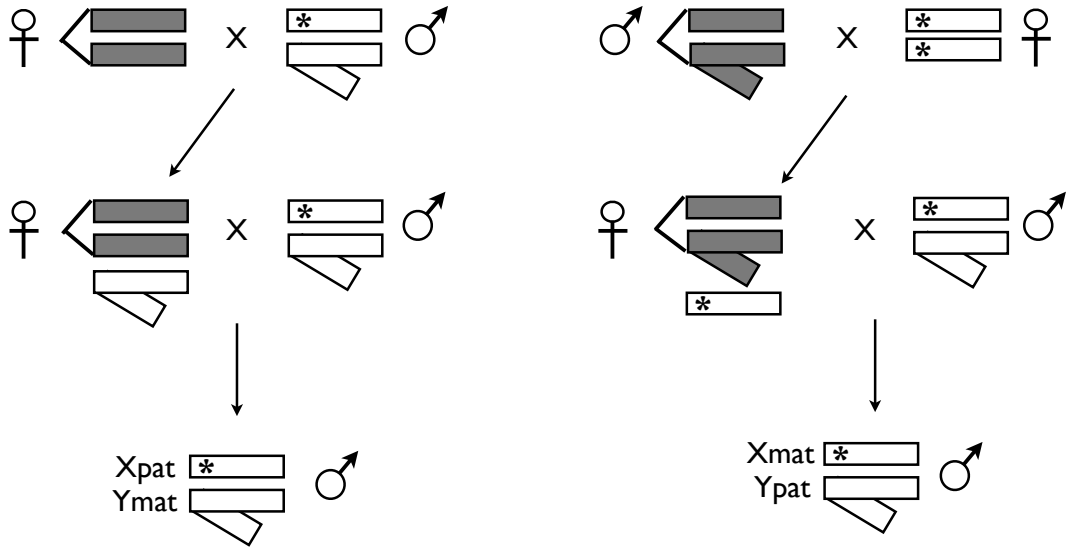
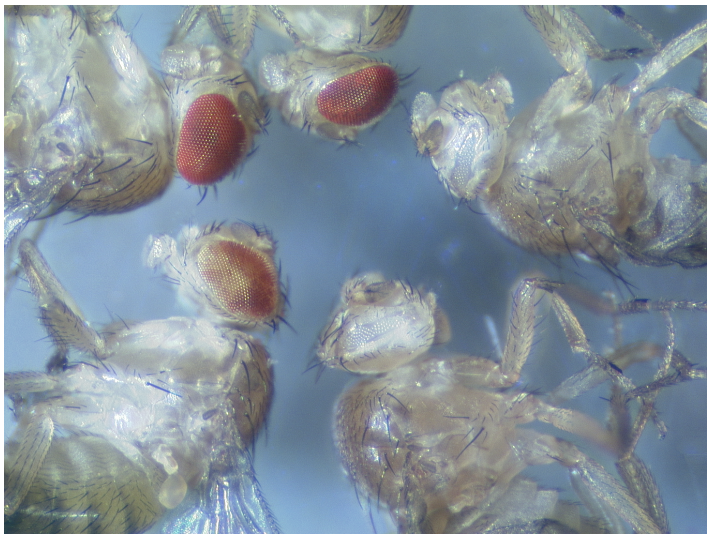


Figure S2.4. Crosses used to generate individuals with reversed sex chromosome parent-of-origin inheritance. In both crosses, maternal genotype was XXY. * indicates the w^{m4h} gene, which was used for PEV assays (see Methods).

A.



B.

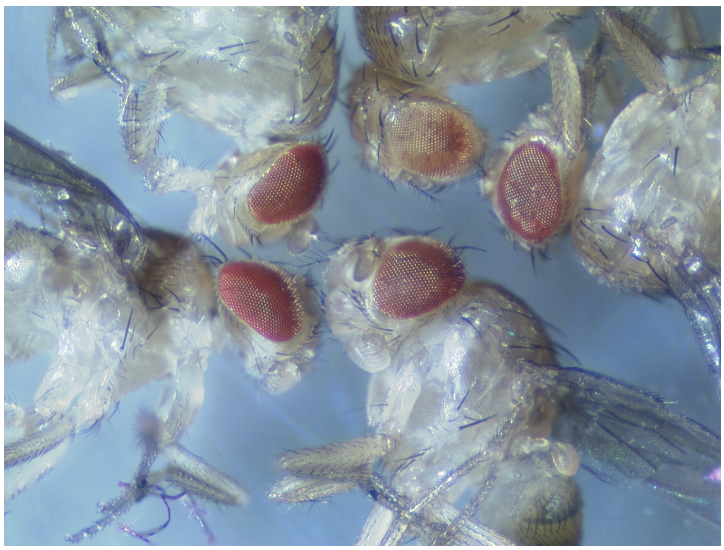


Figure S2.5. PEV differences between males inheriting sex chromosomes as A) normal (Xmat, Ypat), as illustrated in Figure S2.4 right panel, or B) males inheriting reversed parent-of-origin sex chromosomes (Xpat, Ymat), as illustrated in Figure S2.4 left panel. In each case, males had aneuploid XXY mothers.

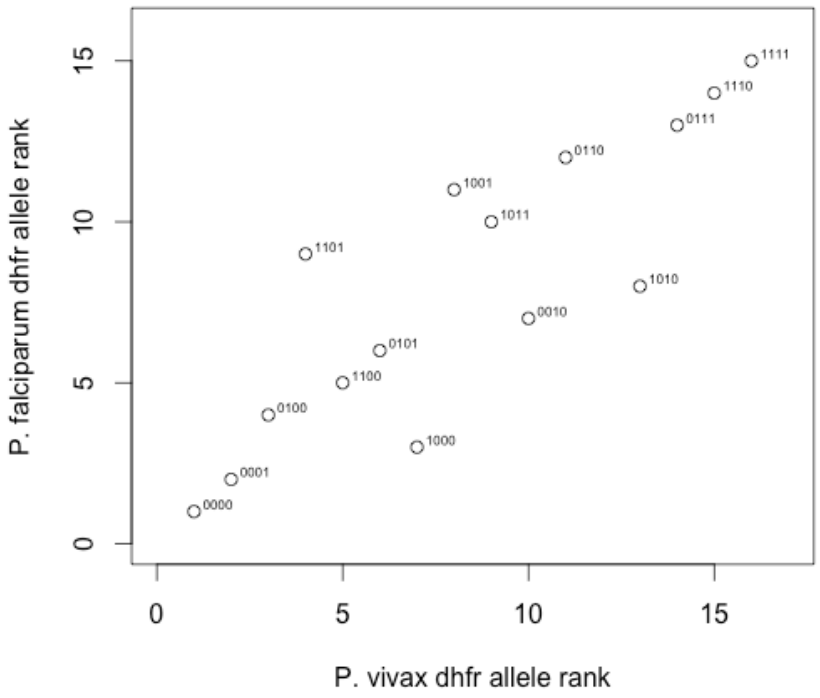


Figure S3.1. Correlations between ranked resistance phenotypes (1 is lowest resistance, 16 is highest resistance) of *P. vivax* and *P. falciparum* DHFR alleles. Allelle 0011 was omitted because this allele in *P. falciparum* DHFR failed to grow.

Figure S3.2. At low concentrations of pyrimethamine, effective population size (Ne) has no effect on time to fixation of an allele since almost all simulations never moved off of the fitness maximum of the wildtype 0000 allele. At higher concentrations of pyrimethamine, Ne has a strong effect on the number of generations required for any particular allele to fix. As expected from population genetics theory, larger Ne increases the efficacy of selection, and reduces the number of generations required for fixation.

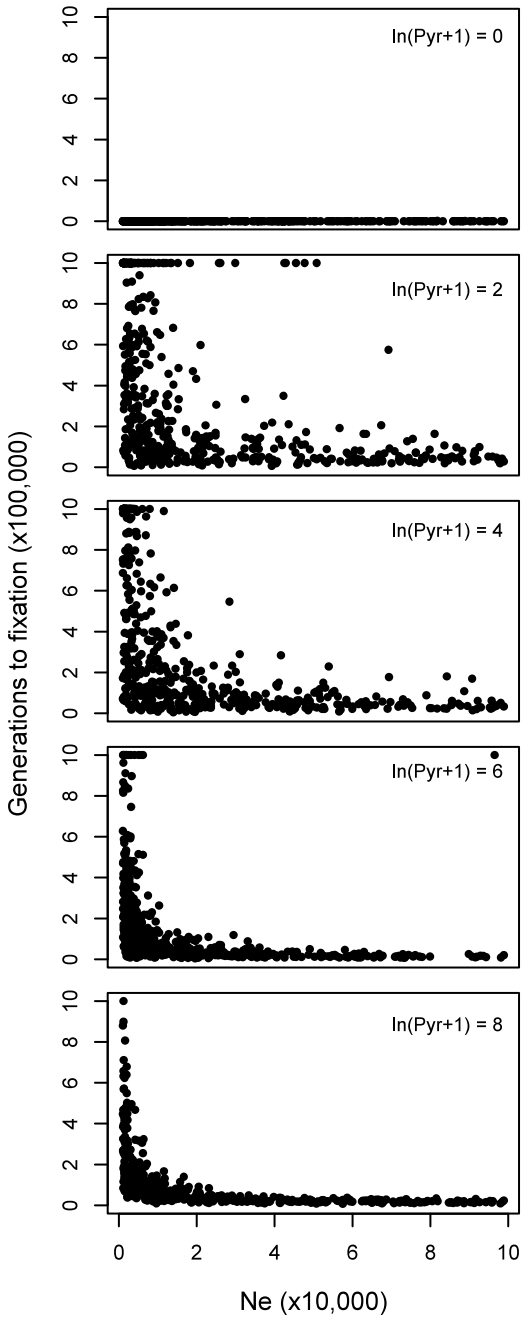


Figure S3.2. (Continued).

Table S3.1: Mean effective population size of simulations which fixed for the highest fitness allele (0000, 1110, 1110, 1110, and 1111 for ln(pyrimethamine+1) concentrations of 0, 2, 4, 6, and 8, respectively), and for all other alleles.

	Highest fitness allele	All other alleles
0	21961	N/A
2	26523	6894
4	24258	2220
6	18737	9576
8	23770	1214

Table S3.2. Comparison of ln(IC50+1) and ranked IC50 values between *S. cerevisiae* and *E. coli* strains carrying the same DHFR alleles. Ranked alleles are indexed from 1 (least resistant) to 16 (most resistant).

Allele	S.c. ln(IC50+1)	S.c. ranked IC50	E.c. ln(IC50+1)	E.c. ranked IC50
0000	0.8104	1	0.3752	1
0001	1.5305	2	2.2106	7
0010	6.5701	10	5.4461	12
0011	7.0782	12	6.4972	15
0100	3.1450	3	0.3791	3
0101	4.6556	6	1.3804	5
0110	6.6353	11	4.2190	8
0111	7.4138	14	6.1458	14
1000	4.7019	7	0.3808	4
1001	5.7108	8	5.4480	11
1010	7.2145	13	4.9754	10
1011	6.3726	9	7.2890	16
1100	4.6333	5	0.3775	2
1101	4.4496	4	1.8584	6
1110	7.5496	15	4.8127	9
1111	7.9816	16	5.4790	13

Table S3.3. Works cited for polymorphic DHFR alleles in *P. vivax*

Allele	Location
S58R/S117N (0110)	Thailand (1), Myanmar (1), Cambodia (1), China (2), India (3), Mauritania (4), Iran (5), Pakistan (6), French Guiana (7), Myanmar (8), Ethiopia (9), Madagascar (10)
S58R/S117N/I173L (0111)	French Guiana (1, 7), Myanmar (11)
S58R (0100)	Thailand (1), India (12), Indonesia (13), Vietnam(14), Vanuatu (14), China (14), East Timor (14)
N50I/S58R/S117N/I173L (1111)	Not found

1. de Pécoulas PE, Tahar R, Ouatas T, Mazabraud A, Basco LK (1998) Sequence variations in the Plasmodium vivax dihydrofolate reductase-thymidylate synthase gene and their relationship with pyrimethamine resistance. *Mol Biochem Parasitol* 92:265–273.

2. Miao M et al. (2010) Different allele prevalence in the dihydrofolate reductase and dihydropteroate synthase genes in Plasmodium vivax populations from China. *Am J Trop Med Hyg* 83:1206–1211.

3. Imwong M et al. (2001) Association of genetic mutations in Plasmodium vivax dhfr with resistance to sulfadoxine-pyrimethamine: geographical and clinical correlates. *Antimicrob Agents Ch* 45:3122–3127.

4. Lekweiry KM et al. (2012) Molecular surveillance of drug-resistant Plasmodium vivax using pvdhfr, pvdhps and pvmdr1 markers in Nouakchott, Mauritania. *J Antimicrob Chemoth* 67:367–374.

5. Afsharpad M, Zakeri S, Pirahmadi S, Djadid ND (2012) Molecular assessment of dhfr/dhps mutations among Plasmodium vivax clinical isolates after introduction of sulfadoxine/pyrimethamine in combination with artesunate in Iran. *Infect Genet Evol* 12:38–44.

6. Khatoon L, Baliraine FN, Bonizzoni M, Malik SA, Yan G (2009) Prevalence of antimalarial drug resistance mutations in Plasmodium vivax and P. falciparum from a malaria-endemic area of Pakistan. *Am J Trop Med Hyg* 81:525–528.

7. Barnadas C et al. (2009) High prevalence and fixation of Plasmodium vivax dhfr/dhps mutations related to sulfadoxine/pyrimethamine resistance in French Guiana. *Am J Trop Med Hyg* 81:19–22.

8. Lu F et al. (2010) Mutations in the antifolate-resistance-associated genes dihydrofolate reductase

and dihydropteroate synthase in Plasmodium vivax isolates from malaria-endemic countries. *Am J Trop Med Hyg* 83:474–479.

9. Schunk M et al. (2006) High prevalence of drug-resistance mutations in Plasmodium falciparum and Plasmodium vivax in southern Ethiopia. *Malaria J* 5:54.

10. Barnadas C et al. (2008) Plasmodium vivax resistance to chloroquine in Madagascar: clinical efficacy and polymorphisms in pvmdr1 and pvcrt-o genes. *Antimicrob Agents Ch* 52:4233–4240.

11. Brega S et al. (2004) Real-time PCR for dihydrofolate reductase gene single-nucleotide polymorphisms in Plasmodium vivax isolates. *Antimicrob Agents Ch* 48:2581–2587.

12. Valecha N et al. (2006) Therapeutic efficacy of chloroquine in Plasmodium vivax from areas with different epidemiological patterns in India and their Pvdhfr gene mutation pattern. *Trans R Soc Trop Med Hyg* 100:831–837.

13. Hastings MD et al. (2005) Novel Plasmodium vivax dhfr alleles from the Indonesian archipelago and Papua New Guinea: Association with pyrimethamine resistance determined by a Saccharomyces cerevisiae expression system. *Antimicrob Agents Ch* 49:733–740.

14. Auliff A et al. (2006) Amino acid mutations in Plasmodium vivax DHFR and DHPS from several geographical regions and susceptibility to antifolate drugs. *Am J Trop Med Hyg* 75:617–621.